A Monte Carlo Approach to Probabilistic Seismic Hazard Analysis in the Aegean Region

Graeme Adrian Weatherill

A Thesis submitted for the degree of Doctor of Philosophy

University of East Anglia

School of Environmental Sciences

May 2009

Abstract

The Aegean is the most seismically active and tectonically complex region in Europe. Damaging earthquakes have occurred here throughout recorded history, often resulting in considerable loss of life. The Monte Carlo method of probabilistic seismic hazard analysis (PSHA) is used to determine the level of ground motion likely to be exceeded in a given time period. Multiple random simulations of seismicity are generated to calculate, directly, the ground motion for a given site.

The three fundamental components of PSHA are considered: seismic source model, magnitude recurrence model and ground motion attenuation model. Initial analysis of the earthquake catalogue indicates that a doubly-truncated Gutenberg-Richter recurrence relation is an appropriate recurrence model for the Aegean. A novel seismic source model is presented, developed by interpretation of Aegean seismotectonics. The K-means cluster analysis algorithm is introduced as new and objective means of partitioning seismicity and seismogenic faults to achieve of source zone delineation. Partitions of the seismicity containing 20 to 30 earthquake clusters emerge as the most appropriate for modelling seismicity in the Aegean. The 27 and 29 cluster K-means source models are integrated into the seismic hazard analysis alongside existing source models. Attenuation models are reviewed, (including European, Greek and global Next Generation Attenuation models) and their suitability for the Aegean region qualitatively and quantitatively assessed.

Seismic hazard maps are produced and site-specific seismic hazard analyses undertaken for 8 selected cities across the Aegean. Epistemic uncertainty is qualitatively assessed by consideration of different source and attenuation models, before being integrated into the PSHA via the Monte Carlo technique. Further extensions to this method (fault and site characterization and aftershock simulation) are presented and their impact on the PSHA assessed. Fault and site characterization appear to have a significant impact on the outcome of the seismic hazard analysis.

Contents

List of Figures

List of Tables

Acknowledgements

Throughout the course of the research for this thesis there have been several people who have provided invaluable guidance and assistance.

I would firstly like to thank my supervisor Dr Paul Burton for his guidance, knowledge and oversight throughout this work.

My colleagues in the Seismic Risk Group, both current and former, have been an invaluable source of support during this research. I am sincerely grateful to Dr Stephen Cole, Dr Antonia Weston, Dr Yebang Xu, James Bayliss and Khaista Rehman. Their presence and assistance have been an enormous reassurance during the difficult periods of this work.

I would also like to acknowledge Dr Susanne Sargeant and Dr Roger Musson of the British Geological Survey for sharing their insights into seismic hazard analysis. I am also grateful that they have given me the opportunity to collaborate with them in other projects that have arisen during the course of this research.

I am very appreciative of several members of the support staff in the School of Environmental Sciences for their assistance. In particular I would like to thank Alan Ovenden (Map Curator) for provision of the mapping material used in this thesis, and Roger Humphrey (Research & Development Technical Manager) and Jenny Stevenson (Fieldstore Manager) for the training they have provided in the use and maintainence of seismic recording equipment.

Additional thanks are extended to Dr Catherine Berge-Thierry for providing a copy of the European Strong Motion database, and also to Dr Christos Papaioannou for sharing his insights into the construction of seismic source models in the Aegean region.

Finally I must express my sincerest gratitude to my family for their support during the course of this research, and throughout my academic career.

Chapter 1

Introduction

1.1 Motivations for Seismic Hazard Research in the Aegean Region

Throughout the history of civilisation, earthquakes have posed a persistent threat to life and property in many regions of the world. In Europe, records of devastating earthquakes and their impact on ancient and modern societies can be found in Italy, Greece, Turkey, the Balkan states and many other countries around the Mediterranean. The earliest of these records date back to the 5th Century B.C. (Papazachos and Papazachou, 1997). During the 20th century, the rapid expansion of cities in the Mediterranean region increased the vulnerability to earthquakes arising from both onshore and offshore active faults. In some countries earthquake risk may be at least partially mitigated by anti-seismic building design and effective codes for building development. Nevertheless the threat to life and property is still significant, and is growing in terms of financial loss.

The eastern Mediterranean provides one of the oldest and most illustrative examples of the interaction between human civilisation and the dynamics of the Earth's crust. A region of geodynamic complexity, a diversity of faulting regimes can be seen around the Aegean. This has resulted in a wide variety of geological hazards that have left indelible marks on both the physical and cultural landscape. In this thesis the Aegean is defined as the area enclosed by the $18^{\circ}E$ and $30^{\circ}E$ meridians, and the $34^{\circ}N$ and $43^{\circ}N$ parallels. This encompasses all of Greece, the Former Yugoslav Republic of Macedonia (FYROM) and Albania, as well as western Turkey and southern Bulgaria.

The greatest step forward in the mitigation of geological hazards is the progress from a fatalistic attitude to a preparatory and proactive one. Despite political upheaval for much of the early 20th century, the recognition by Greek, Turkish and other Balkan governments of the need for earthquake mitigation has been a progressive step forward. As many countries in Eastern Europe have gained economic strength over the past 30 to 40

years, this has given rise to extensive urban development, resulting in greater economic vulnerability. Where a major earthquake may severely damage the civil infrastructure of a city, the economic impact may be highly significant nationally. For example, the 1999 Izmit earthquake resulted in an estimated total economic loss of US$ 15 - 20 billion; approximately 7 % of Turkey's GDP as of 1999 (Bendimerad *et al.*, 2000).

The human and economic cost of earthquakes in the Aegean region has driven research in earthquake science in this part of the world. The consequences of large earthquakes across the globe are a primary motivation for understanding seismic hazard. This thesis is intended to make a contribution to the development of a consistent and transparent procedure for the analysis of this hazard. Particular consideration is given to the appraisal of seismic hazard in the context of Aegean seismotectonics. The methods developed in this work can be, and already have been, applied in other parts of the world.

1.2 Seismic Hazard Analysis

Underpinning all the currently viable means of earthquake mitigation across the globe, are quantitative analyses of seismic hazard at a location. This is commonly defined as the level of peak ground motion occurring at a location, with a P % probability of being exceeded in a period of T years. This information is an important element of both economic and engineering means of mitigating earthquake effects. Knowledge of the seismic hazard across an urban area is also essential for effective underwriting and implementation of earthquake insurance. In combination with databases of building stock and estimated value, this allows governments and insurers to assess the likely losses in the event of an earthquake. This is defined as the seismic risk:

$$Seismic\,Risk(\$) = Seismic\,Hazard \times Vulnerability \times Value(\$) \tag{1.1}$$

Seismic hazard analyses are widely used in civil engineering applications. Improvements in seismic resistance of both large engineered structures (e.g. bridges, dams, power generation etc.) and domestic properties have resulted in a reduction in loss of life following major earthquakes; at least in areas where they can be rigorously enforced. However, anti-seismic design comes at an increased cost. It is therefore essential for engineers to understand the degree of strong shaking that their structures may be likely to experience within their expected design life.

The volume of published literature on the topic of seismic hazard analysis that has been developed over the past half century is considerable. The definition of seismic hazard, in addition to the context of its application, requires that it be quantified in a probabilistic sense. Probabilistic Seismic Hazard Analysis (PSHA) has therefore become the standard approach to assessing the level of seismic hazard, given the available information. The most common technique for this purpose is the Cornell (1968) - McGuire (1976) method

of PSHA. For a site located within distances R_i of i seismogenic sources, each capable of producing earthquakes in the magnitude range $M_{MIN} \leq M \leq M_{MAXi}$, and assuming random (aleatory) variability in the ground motion attenuation described by ε standard deviations from the median, the rate at which ground motion Y exceeds a specific level Y_0 is given by (Cornell, 1968; Kramer, 1996; Abrahamson, 2000; McGuire, 1995):

$$\sum_i \lambda_i \int_{r=0}^{\infty} \int_{m=M_{MIN}}^{M_{MAXi}} \int_{\varepsilon=\varepsilon_{MIN}}^{\varepsilon_{MAX}} f_M(m)\, f_R(r)\, f_\varepsilon(\varepsilon)\, P[Y > Y_0 | m, r, \varepsilon] dm dr d\varepsilon \tag{1.2}$$

where λ_i is the rate of occurrence of earthquakes with magnitude greater than M_{MIN} from the i^{th} source. These integrals are evaluated numerically. Assuming that ground motion exceeding level Y_0 within a time period of T years can be expressed as a Poisson process, seismic hazard is determined by:

$$P(Y > Y_0, T) = 1 - e^{-\lambda T} \tag{1.3}$$

The functions $f_R(t)$, $f_M(m)$, and $f_\varepsilon(\varepsilon)$ are defined in the PSHA process. They are characteristics of the source model, magnitude-recurrence model and attenuation model respectively. This is illustrated in Figure 1.1 (Reiter, 1990; Kramer, 1996). Each step of the PSHA process will be described in detail in the following chapters.

Whilst widely used, the Cornell-McGuire method in the form illustrated here suffers from several problems that have, at times, attracted criticism (Krinitzsky, 2002; Klügel, 2005). Amongst the criticisms are: 1) difficulty in incorporating uncertainties in the parameters of the recurrence relation, 2) a dependence on a Poissonian assumption of earthquake behaviour and 3) difficulty in defining design earthquakes. Furthermore, the Cornell-McGuire method is often perceived as lacking transparency to the final user, failing to elucidate the relative contribution to the seismic hazard from different sources and magnitudes. Many of these problems are addressed in the analysis of epistemic uncertainty (uncertainty arising from insufficient knowledge of the process), but implementation of such analysis can become analytically and computationally unwieldy.

To address some of these problems a different but equally compatible procedure has been developed. This is a Monte Carlo procedure, which utilises multiple random simulations of seismicity for the purposes of seismic hazard analysis. An early framework for the Monte Carlo procedure was developed for PSHA in Israel by Shapira (1983), and later in Hawaii (Johnson and Koyanagi, 1988). Unfortunately, computational efficiency was such that the Monte Carlo method was inefficient at the time of this initial development. After remaining out of favour for more then 10 years, two new Monte Carlo routines were developed for PSHA. These routines differed slightly in their approaches to seismicity simulation. The first used random re-sampling (with replacement) of observed seismicity to generate synthetic earthquake catalogues (Ebel and Kafka, 1999). The second method simulated seismicity by random sampling of the source, recurrence and at-

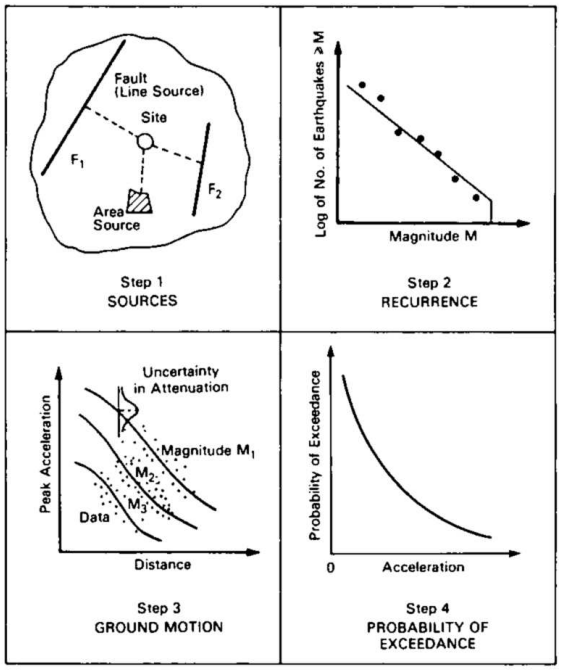

Figure 1.1: Illustration of the basic elements of a probabilistic seismic hazard analysis (TERA, 1980)

tenuation models, similar to the schematic illustrated in Figure 1.1. This approach has since been applied to seismic hazard analysis in the Balkans (Musson, 1999b), the United Kingdom (Musson, 2000; Musson and Sargeant, 2007) and Switzerland (Giardini *et al.*, 2004; Wiemer *et al.*, 2009). The two different procedures by which synthetic earthquake catalogues of duration T_0 are generated and used to calculate hazard are shown in Figure 1.2.

Both procedures for PSHA using Monte Carlo simulation produce results that are compatible with the Cornell-McGuire method. However, they do so at a substantially greater computational cost. This accounts, at least partially, for why the procedure did not prove popular at the time of its initial development, but has become more widely used in the last few years. Monte Carlo PSHA can now be run on a "normal" desktop or laptop computer within a reasonable period of time (less than 10 seconds for one site).

In addition to being compatible with the Cornell-McGuire method of PSHA, the Monte Carlo approach has many advantages. The greatest advantage is in its ability to handle uncertainty in the input parameters of particular models. Input parameters are fixed in the Cornell-McGuire method, and uncertainty explored by sensitivity analysis or epistemic methods, such as logic tree analysis (Coppersmith and Youngs, 1986). Monte Carlo allows for parameters to be sampled from probability distributions of observed mean and variance (Ebel and Kafka, 1999; Musson, 2000), thus incorporating uncertainty on these parameters into the probabilistic calculation. Similarly, different models of earthquake behaviour can be explored, without the complexity that accompanies the need

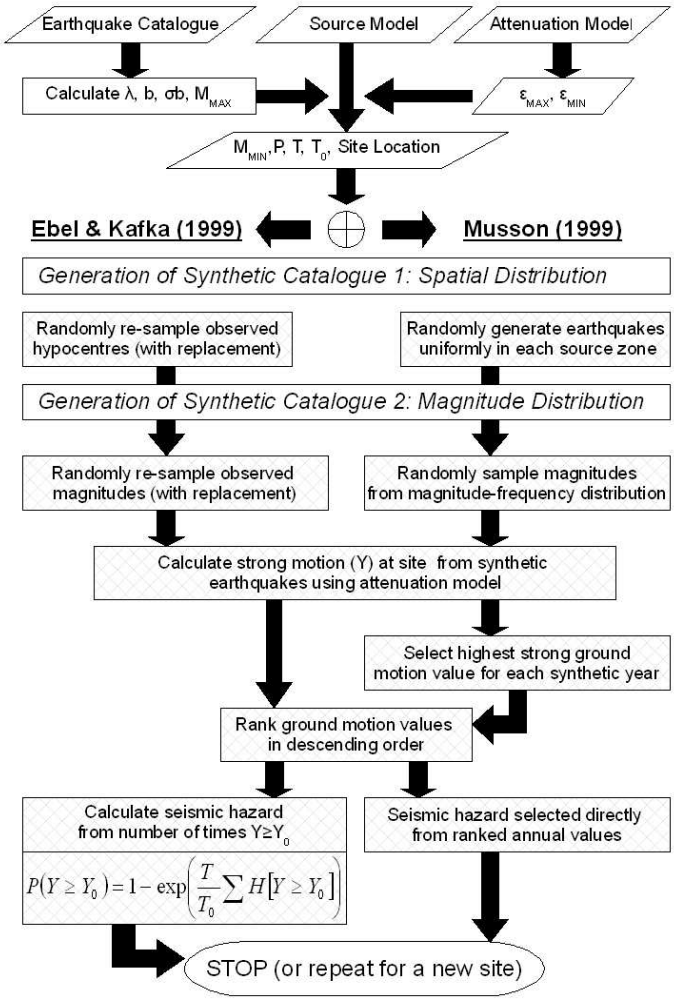

Figure 1.2: Flowchart of the Ebel and Kafka (1999) (left branch) and the Musson (1999b) (right branch) of the Monte Carlo procedure

to describe them as an integral function in the manner of equation 1.2. This makes the Monte Carlo approach more transparent and conceptually straightforward than that of Cornell-McGuire. Also, by making an explicit record of the influence of each synthetic earthquake on hazard at a site, the Monte Carlo procedure can clearly identify design earthquakes (Musson, 1999a). This can make it a powerful tool for engineering design as well as seismic risk analysis.

1.3 Thesis Objective

The objective of this work is to implement a full seismic hazard analysis for the entire Aegean region using the Monte Carlo procedure. This includes a comprehensive and transparent assessment of the sources of uncertainty that have been endemic, but not always explicit, in previous seismic hazard analysis in this region. It is also the intention of this work, to converge toward a more objective and replicable procedure for seismic

hazard analysis.

The primary focus of the project is comprehensive assessment of the spatial variation in seismic hazard by means of a hazard map. As such, it is expected that one outcome of this thesis will be a "recommended" seismic hazard map for a given probability level in the Aegean region. This map can form the basis for seismic zoning in Greece, Albania and FYROM, for the purposes of employing building codes for anti-seismic design. A map of this nature can be used for guidance in assessing regional variation in implementation of Eurocode 8 (the current European standard building code).

Given the emphasis placed on spatial data distribution, an important focus of the initial chapters of this thesis is assessment of the variability in data quality across the Aegean. This will include issues such as earthquake catalogue completeness and quality, as well as precise identification of seismogenic sources. Unfortunately, experience in working with such data from this region indicates that regional variability in data quality is an enormous problem. No single agency or geological survey operates across all the Aegean countries studied here. This has meant that much of the research into the location and nature of specific faults is undertaken by individual Universities and research institutions. The result is observational scientific disparity, with different institutions producing contrasting interpretations of fault location and behaviour within the same region. Furthermore, seismotectonic maps and databases dislocate across political borders. This work refrains from making any judgement as to which interpretation or database is correct. Instead, a key focus will be the illustration of how a seismic hazard analysis can be performed on a more information-limited, but spatially homogenous, dataset.

The demands of a site-specific seismic hazard analysis for engineering applications are different from those of spatial assessment of seismic hazard. In practice, more detailed information regarding the seismic response of the site, and the physical properties of nearby seismogenic faults, is required. Such detailed research is beyond the scope of this thesis. However, an illustration of how the Monte Carlo method can be applied to, and fulfil many of the demands of, site specific hazard analysis is a secondary objective of this investigation.

The software used to undertake the Monte Carlo seismic hazard analysis here is mostly original, with existing routines referenced where necessary. It is hoped that this will form the initial stage of an endeavour to create an accessible program to compute seismic hazard using stochastic techniques.

1.4 Thesis Outline

This work is organised into seven chapters, which formally discuss the Monte Carlo seismic hazard analysis process:

Chapter 2 introduces the Aegean earthquake catalogue, which is then analysed in terms of spatial and temporal completeness as well as magnitude recurrence. Discussion and justification of the preferred magnitude-recurrence model, and the upper and lower magnitude limits for the seismic hazard analysis, are also given in this chapter.

Chapter 3 compares existing models of seismogenic sources in the Aegean that have been developed for the purposes of seismic hazard mapping. Discussion of uniform source zone models and zone-free source characterisation can be found in this chapter. Consideration is then given to the existing uniform source models that have been applied to previous seismic hazard analyses in the Aegean region. The chapter concludes with the introduction of a new seismic source model and discussion of the data used and interpreted in its creation. This model represents this author's interpretation of Aegean seismotectonics in the context of delineating uniform seismic source zones. For the purposes of transparency a full zone-by-zone discussion can be found in Appendix A.

Chapter 4 presents a novel approach to objective delineation of seismic sources for seismic hazard analysis using the K-means cluster analysis algorithm. A description and justification of the method is given. It is then applied to catalogues of Aegean seismicity and fault ruptures. Partitions are then analysed in the context of seismotectonic variation in the Aegean. The results of this chapter are published in Weatherill and Burton (2009), and an application of this technique to seismic hazard analysis in Java, Indonesia, can also be found in Burton *et al.* (2008).

Chapter 5 details the decisions and considerations that are needed to select an appropriate empirical ground motion attenuation model. Candidate relations are presented and preliminary selection is made, based on a qualitative selection criteria. Those relations that are not rejected using qualitative criteria are then quantitatively compared with observed strong motion records from the Aegean region. Weightings for epistemic uncertainty analysis are apportioned on the basis of their fit to the observed strong motion data.

With the inputs for the analysis in place, the remaining chapters present the results of the seismic hazard assessment in the Aegean using the Monte Carlo method:

Chapter 6 presents the seismic hazard maps created by "single model" hazard analysis, i.e. analysis using a specific combination of source and attenuation model. The impacts of particular source and attenuation models are compared in both a hazard mapping and site-specific context. The Monte Carlo method is also used to obtain design earthquakes via a disaggregation technique (Musson, 1999a).

Chapter 7 illustrates several novel extensions to the Monte Carlo SHA process. These include integration of epistemic uncertainty into the seismic hazard assessment, incorporation of intermediate depth earthquakes, aftershocks, simple fault models and variation

in site condition.

The conclusions of this thesis are presented in chapter 8 and the highlights of the research identified. Also shown is the preferred seismic hazard map (and models), given the material presented in this thesis.

Chapter 2

Observed Seismicity and its Input for Seismic Hazard Analysis

The historical catalogue of earthquakes is the most fundamental component in any seismic hazard analysis. It is from such catalogues that parameters describing earthquake behaviour (e.g. recurrence, maximum magnitude) are determined. In places where additional seismotectonic information is available (e.g. geodetic strain, average slip rate), this can be used to refine understanding of the earthquake behaviour in the region. For most applications, however, the primary source of information is the earthquake catalogue.

For the stochastic seismic hazard analysis applied to the Aegean region, this chapter will outline the earthquake catalogues used for input. It has been fortuitous that, at present, several good quality earthquake catalogues are available. These have been assembled from both historical and instrumental records of Aegean earthquakes, many of which have been published in various journals throughout the last 30 to 40 years. The most up-to-date catalogues are discussed here.

Although the earthquake catalogue may be one of the primary inputs for seismic hazard analysis, there are many pieces of information that must be extracted for use in seismic hazard analysis. This includes completeness magnitude (the magnitude below which it cannot be assumed that all occurring events are recorded in the catalogue), magnitude-frequency recurrence, seismicity rate, maximum magnitude and threshold or minimum magnitude for PSHA. Analyses of these parameters and their relation to Aegean seismotectonics are also presented here. Discussion of current debate and controversy regarding some of these methods and models is also given.

2.1 The Earthquake Catalogues

Three catalogues of earthquakes are used in this work. The first is a catalogue of large historical earthquakes, spanning the period 550 B.C. to 1900 A.D (Papazachos and Papazachou, 1997). This catalogue covers the area 18° E to 30° E and 34° N to 43° N. The largest earthquake in this catalogue is the 21 July A.D. 365 Gortyna (South Aegean) event ($M_W \approx 8.3$). The catalogue can be considered complete above M_W 6.5 for the period 1650 A.D. to present, and complete above M_W 6 for the period 1800 A.D. to present. It contains no events smaller than M_W 6.0. No focal depth information is available for most of the earthquakes in the catalogue. Where earthquakes are recognised as being intermediate depth or deep they have been assigned arbitrary depths, usually 75 km (intermediate) or 150 km (deep).

The second catalogue considered is the 20th Century catalogue of Burton *et al.* (2004a). This catalogue covers the same area as the Papazachos and Papazachou (1997) catalogue and spans the period 1901 A.D. to 1999 A.D. This can be broken down into three sub-periods: 1901 - 1963, 1964 - 1998 and 1999. Earthquakes from the first sub-period are taken from two catalogues: Makropoulos and Burton (1981) and Shebalin *et al.* (1974). Earthquakes in this sub-period were rendered onto the Surface-Wave Magnitude (M_S) scale. In the original (Makropoulos and Burton, 1981) earthquake hypocentres were directly relocated for events prior to the International Seismological Centre period (1964 onwards). Similarly, surface wave magnitudes were re-measured directly, rather than rendered from local magnitude scales. For the final catalogue, moment magnitudes were then calculated using the equations of Papazachos and Papazachou (1997).

$$M_W = 0.804\, M_S + 1.28 \quad 5.3 \leq M_S \tag{2.1}$$

$$M_W = 0.56\, M_S + 2.66 \quad 4.2 \leq M_S \leq 6.0 \tag{2.2}$$

The smallest recorded earthquakes in this sub-period are M_W 4.9, although the catalogue is only complete to M_W 5.2. The largest earthquake is the 11 August 1903 Kythera earthquake (M_W 7.7). This sub-period contributes 1,285 earthquakes. Estimates of focal depth are rounded to the nearest kilometre.

The second sub-period of the Burton *et al.* (2004a) catalogue spans the period 1964 - 1998 A.D.; contributing 2,390 earthquakes in the magnitude range $3.2 \leq M_W \leq 6.9$. This catalogue is taken from the standard data of the International Seismological Centre (Engdahl *et al.*, 1998). These earthquakes are reported in either moment magnitude or body-wave magnitude (m_b). The latter are homogenised into M_W using the relation of Papazachos and Papazachou (1997):

$$M_W = 1.28 m_b - 1.12 \qquad \text{for} \qquad 4.8 \leq m_b \leq 6.0 \tag{2.3}$$

Body-wave magnitudes are converted to M_W using the following relation:

$$M_S = (1.8782 \pm 0.0222)\, m_b - (4.6046 \pm 0.1102) \tag{2.4}$$

Focal depths for this sub-period are given to the nearest 0.1 km, though errors in the focal depth are likely to be on the order of at least 1 - 2 km.

The third sub-period of the Burton *et al.* (2004a) catalogue covers only the year 1999. Despite the brevity relative to the full catalogue length, this sub-period provides a procedure for continuation when ISC data are unavailable. Earthquakes in this period come from the National Observatory of Athens Online catalogue (NOA, 2007 - www.gein.noa.gr/services/info-en.html). Magnitudes in this catalogue are given NOA local magnitude, which is converted to M_W using the relation of Margaris and Papazachos (1999):

$$M_W = M_L + 0.43 \tag{2.5}$$

This is converted into surface wave magnitude M_S using the relation given by the NOA Monthly Bulletin:

$$M_S = 1.70\,(\pm 0.05)\, M_L - 3.59\,(\pm 0.22) \tag{2.6}$$

The 17 August 1999 Izmit earthquake (M_W 7.6) is added to this catalogue from the Harvard CMT catalogue, as well as five significant aftershocks of this event. The combined NOA and Harvard CMT catalogues contribute a further 2,502 earthquakes in the range $2.1 \leq M_W \leq 7.6$. Focal depths are given to the nearest 0.1 km, though the errors on these depth determinations are likely to be much greater greater.

The Burton *et al.* (2004a) catalogue is truncated at M_W 4.0. This is to remove microearthquakes that are inconsistently reported, whilst retaining events of possible engineering significance. The resultant catalogue for the period 1901 - 1999 therefore contains 5,198 events in the magnitude range $4.0 \leq M_W \leq 7.7$, with focal depths in the range 1 - 215 km.

A third catalogue is added to include events in the period 2000 A.D. to 2005 A.D. This catalogue is also taken from the NOA catalogue, which spans the area $18.5^\circ E$ to $30^\circ E$ and $33^\circ N$ to $42^\circ N$. The earthquakes are reported in terms of NOA local magnitude. These are converted to M_W and M_S using the relations in equations 2.5 and 2.6 respectively. This catalogue contains a further 3,782 events in the magnitude range $M_W \geq 4$. The largest earthquake in this catalogue is the 22 January 2002 Karpathos earthquake (M_W 6.5). The magnitude of completeness for the 1999 - 2005 period is below M_W 4.0.

2.2 Completeness

The combination of catalogues that now spans the time period 550 B.C. to 2005 A.D. contains 9,407 events with moment magnitude M_W 4.0. As is noted in Section 2.1, variation in quality of seismic recording and data availability, in addition to social factors such as the distribution of population, mean that it cannot be assumed all earthquakes above this magnitude are present in the catalogue. Detection of events can depend on the location and quality of recording apparatus (including macroseismic records); hence there is often a bias against smaller shocks in the earlier parts of the catalogue. It is known that the catalogue is complete for events $M_W \geq 4.0$ during the NOA period (1999 - 2005). For the 20th century, however, completeness must be analysed using statistical techniques.

2.2.1 The Stepp (1971) Method of Completeness Analysis

The Stepp (1971) method of completeness analysis operates on the assumption that the number of earthquakes (k) within a unit time interval can be modelled by a Poisson distribution:

$$f(k;\lambda) = \frac{\lambda^k e^{-\lambda}}{k!} \tag{2.7}$$

where λ is the maximum likelihood estimate of the mean number of earthquakes per unit time interval from the set of n time intervals ($k_1, k_2 \ldots k_n$), defined as:

$$\lambda = \frac{1}{n}\sum_{i=1}^{n} k_i \tag{2.8}$$

with variance $\sigma_\lambda^2 = \lambda$. Hence for a sample of length T, the standard deviation is defined as:

$$\sigma_\lambda = \sqrt{\lambda/T} \tag{2.9}$$

Assuming stationarity, the mean and variance should remain constant; hence σ_λ behaves as $1/\sqrt{T}$ in the interval for which the mean rate of occurrence remains constant. If the observed σ_λ departs from the $1/\sqrt{T}$ behaviour then either the interval is not long enough to give a robust estimate of the mean, or the interval includes periods for which the catalogue is incomplete. Analysis is chosen for five magnitude classes: $4.5 \leq M_W < 5$; $5 \leq M_W < 5.5$; $5.5 \leq M_W < 6$; $6 \leq M_W < 6.5$ and $M_W \geq 6.5$. The plot of σ_λ with time can be seen in Figure 2.1 and the temporal variation in completeness shown in Table 2.1.

The Stepp (1971) method of analysis clearly reflects some of the key variations in temporal completeness of the earthquake catalogue. For the $4.5 \leq M_W < 5$ class, deviation from the $1/\sqrt{T}$ line occurs at a duration of 40 years. This corresponds to the length of

Table 2.1: Completeness Periods Estimated from Stepp (1971) Analysis

Time Period	Number of Years	M_C
2005 - 1999	7	4.0
2005 - 1965	40	4.5
2005 - 1900	105	5.0
2005 - 1800	205	6.0
2005 - 1700	305	6.5

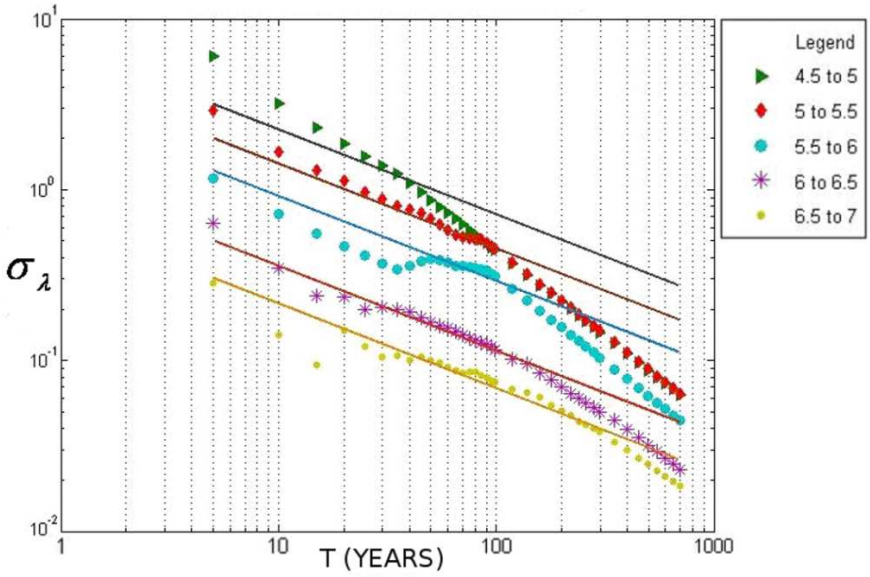

Figure 2.1: Analysis of temporal variation in completeness for the full Aegean catalogue (550 B.C. to 2005 A.D.) using the Stepp (1971) method

the ISC catalogue, which begins in 1964. Both the $5 \leq M_W < 5.5$ and $5.5 \leq M_W < 6$ classes appear complete for a period of approximately 110 years. This suggests that a completeness magnitude (M_C) of M_W 5.0 is reasonable for the entire 20th century catalogue, though this may be an underestimate as shall be seen in the next section. The appending of the Burton *et al.* (2004a) and Papazachos and Papazachou (1997) catalogues, produces a sudden leap in completeness from M_W 5 to M_W 6 at the end of the 19th century. This is due to the higher cut-off magnitude of the Papazachos and Papazachou (1997) catalogue.

The Stepp (1971) method provides a useful overview of temporal variation in completeness, but the precise value of M_C depends on the magnitude classes used. The 20th century M_C estimate of M_W 5.0 may be lower or higher than the true M_C because smaller variation is hidden within the class width. To constrain M_C more accurately other methods are used. Each of these methods allows for analysis of M_C across a much narrower magnitude bound, typically 0.1 magnitude units (Wiemer and Wyss, 2000; Wiemer, 2001; Wössner and Wiemer, 2005). They also require the fit of a cumulative magnitude frequency distribution, mostly the Gutenberg and Richter (1944) distribution.

2.2.2 Maximum Curvature

The first method of analysis is the Maximum Curvature method. This is by far the simplest method and one of the most robust (Wössner and Wiemer, 2005). Since it is assumed that the cumulative magnitude-frequency follow a power law distribution then the absolute number of events for each magnitude range should increase at lower magnitudes. Below the magnitude of completeness the number of events begins to decrease. From this simple assumption the magnitude of completeness is the magnitude that occurs most frequently.

This is arguably an oversimplification, and for some catalogues may lead to problems of underestimation or overestimation of M_C. It assumes that below M_C the number of events decreases, yet the absolute number of events doesn't necessarily have to decrease to deviate from the power-law behaviour. It may be the case that there is an increasing deficit of smaller earthquakes within the catalogue. This will produce a gradual deviation from the expected power-law, without necessarily producing a peak in absolute number of events of a particular magnitude, thus lowering M_C. Similarly, for shorter catalogues a transient feature such as a swarm of small-moderate events may produce an anomalously high number of events above M_C. This will artificially raise M_C. Despite these problems, this technique remains one of the robust methods for estimation of M_C.

Temporal variation in completeness, and its uncertainty, using the Maximum curvature method and an alternative Entire Magnitude Range (EMR) method (Wössner and Wiemer, 2005) is shown in Figure 2.2. The EMR method fits, via a maximum likelihood approach, a cumulative normal distribution to the cumulative number of events below M_C and the Gutenberg-Richter relation to events above it. M_C is the magnitude for which the log-likelihood function is maximised.

For both the Maximum Curvature and EMR methods shown in Figure 2.2, temporal variation in completeness is determined using a moving time window approach. A window of 10 years is used to determine completeness, which is then shifted in 1-year increments. Hence at no time does the catalogue duration exceed 10 years. An alternative approach would be an "apparent M_C" analysis, where the catalogue duration increases incrementally by one year, as will be shown in Section 2.2.4.

For the early 20th century period, the maximum curvature method suggests $M_C = 5.2 \pm 0.1$. The completeness magnitude does rise during the 1920 - 1950 period, before then dropping to M_W 4.8 at the beginning of the ISC period. A similar trend is borne out when using the EMR method, albeit with a higher M_C of M_W 5.5 estimated for the early 20th century. These results suggest that the M_C 5.0 determined from the Stepp (1971) analysis is not entirely accurate and that the completeness magnitude for the 20th century lies somewhere in the $5.0 \leq M_W \leq 5.5$ range.

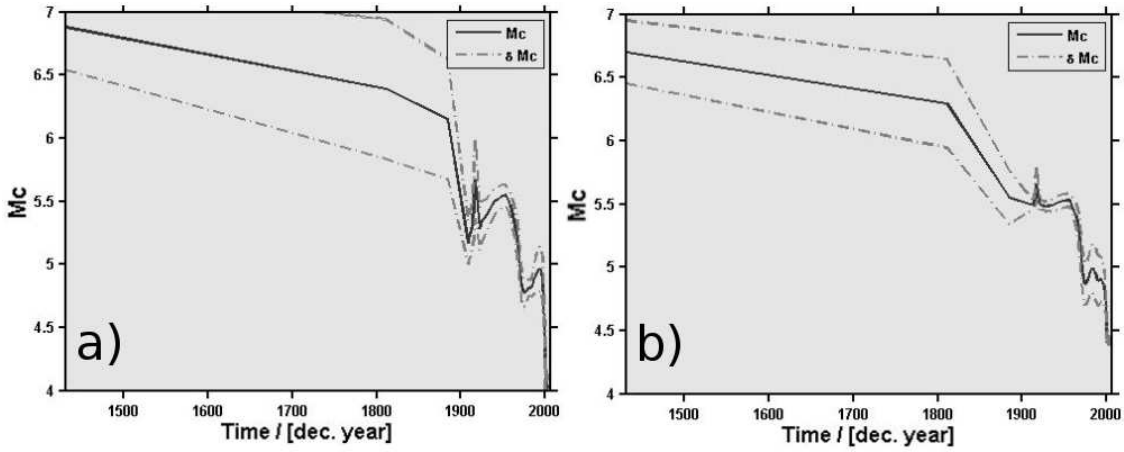

Figure 2.2: Temporal variation in completeness using the ZMAP methodology (Wössner and Wiemer, 2005). a) Maximum Curvature Method, b) Entire Magnitude range method

2.2.3 Goodness of Fit Test

This method of estimating M_C tests the fit of the Gutenberg-Richter relation for the cumulative magnitude-frequency distribution at incrementally increasing cut-off magnitudes (Wiemer and Wyss, 2000). The lowest magnitude to achieve a fit of P % is the magnitude of completeness, where residual R is defined as:

$$R = 100 - 100 \left(\frac{\sum_{M_i}^{M_{MAX}} |B_i - S_i|}{\sum_i B_i} \right) \tag{2.10}$$

where B_i and S_i are the observed and predicted cumulative number of events in each magnitude bin (here 0.1 M_W). The estimate of M_C may depend on the probability level of fit, which in turn is influenced by the total number of events in the catalogues. The 90 % probability level is used here. Whether the goodness of fit reaches the 90 % level may depend on the cumulative frequency-magnitude distribution itself. If there is non-linear behaviour of the log cumulative magnitude-frequency at higher magnitudes, this will reduce the goodness of fit. As such, shorter catalogues with transient features seismicity features produce poor results and artificially high M_C when using this method.

2.2.4 b-value Stability

This is another method that assigns M_C based on the fit of the Gutenberg-Richter relation. If it is assumed that for $M_{CUTOFF} < M_C$, b-values will be lower owing to the inclusion of incomplete magnitudes, and will increase as M_{CUTOFF} tends towards M_C. As M_{CUTOFF} increases above M_C, the b-value will stabilise. M_C is then determined as the lowest magnitude at which the difference in b-value for successive M_{CUTOFF} (Δb) is below a pre-defined threshold. Using the criteria of Wössner and Wiemer (2005), this

threshold is defined as:

$$\Delta b = |b_{ave} - b| \leq \delta b \tag{2.11}$$

Where δb is the uncertainty on b-value determined using the Shi and Bolt (1982) criterion. b_{ave} is defined as:

$$b_{ave} = \sum_{M_{CO}=1.5}^{2} b\left(M_{CO}\right)/5 \tag{2.12}$$

This is the average b-value over five successive cut-off magnitudes in a 0.5 magnitude unit bin range. As with the goodness of fit test, the b-value stability is affected by the fit of the Gutenberg-Richter relation. Non-linearity at higher magnitudes may have the impact of altering the b-value stability, though this is reduced when using the Maximum Likelihood method of estimating b-value (Aki, 1965).

The time variation in completeness for the 20th century is tested here by calculating completeness for catalogues of incrementally decreasing length. Starting with the period 1901 - 2005, M_C is calculated using the maximum curvature, goodness of fit and b-value stability methods. The catalogue is then shortened by 1 year (1900 - 2005), and the process repeated. Analysis of time-varying M_C by estimation of M_C for catalogues of incrementally increasing length is shown in Figure 2.3 The result is a plot of variation in apparent completeness with time.

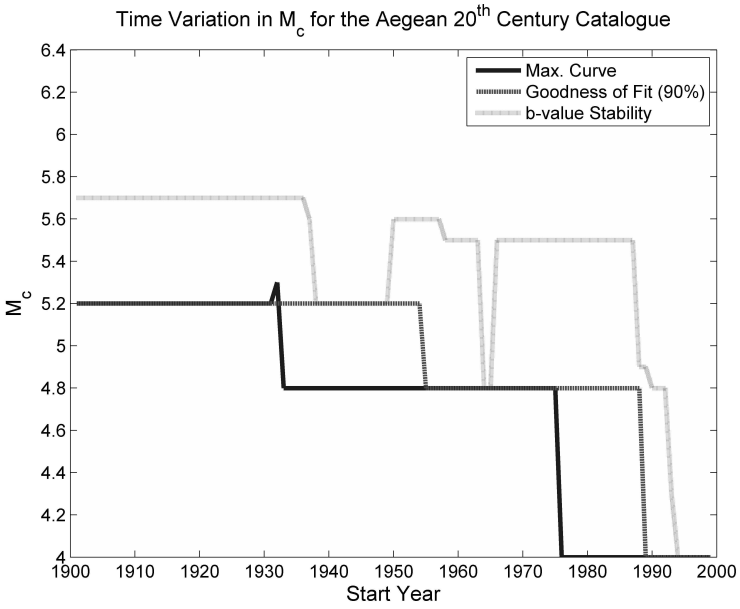

Figure 2.3: Temporal variation in apparent M_C using incrementally increasing catalogue length

Figure 2.3 illustrates that both Maximum Curvature and 90 % Goodness of Fit agree on a completeness magnitude of M_W 5.2 for the 20th century period. The b-value stability estimate is considerably higher (M_W 5.7) for the early 20th century and shows much

more variability. M_C drops to M_W 4.8 roughly in the middle of the 20th century and then again to M_W 4 (or lower) in the last quarter of the century. There is reasonable agreement in M_C between the Maximum Curvature and 90 % Goodness of Fit estimates; the differences being the time of the change. Unsurprisingly, "apparent" M_C decreases typically 10 to 15 years in advance of a change in the catalogue structure (1950 for the ISC data, 1985 for the NOA data). This represents the difference between "apparent" M_C, which comes from including later earthquakes in the analysis, and time window Mc which only considers earthquakes in a moving time interval (10 years in 2.2).

Figures 2.1 - 2.3 illustrate the observation that for the whole Aegean catalogue, $M_C \approx 5.2$ for the 20th century period. A clear drop in M_C to M_W 4.8 is visible from 1964 onwards, which represents the establishment of the International Seismological Centre, and a further drop to M_W 4 is visible around 1998/1999 with the use of the NOA catalogue. The estimates of M_C are subject to error and not all of the methods used to analyse it agree, as is visible in Figure 2.3. Generally the Maximum Curvature method appears to give the most robust estimate of M_C, although the Entire Magnitude Range method (Wössner and Wiemer, 2005) performs well, albeit at considerably greater computational cost. Completeness magnitude can also display enormous spatial variation with the highest quality networks operating in highly active or metropolitan areas. Spatial analysis of M_C using an automatic approach is greatly affected by short-term of transient features in the earthquakes catalogue.

2.3 Removal of Non-Poissonian Events (Declustering)

For the purposes of time-independent seismic hazard analysis, non-Poissonian events have to be removed from the catalogue. This is necessary since a fundamental assumption of the PSHA procedure is that of Poissonian seismicity. Inclusion of non-Poissonian events biases the estimation of magnitude recurrence parameters. The removal of non-Poissonian events by algorithmic means is a complex process that readily leads to misidentification of aftershocks and foreshocks. Several algorithms are in widespread usage (Gardner and Knopoff, 1974; Reasenberg, 1985; Musson, 1999b). Of these, the most common is the algorithm of Reasenberg (1985), which is used in this study. Most declustering algorithms operate by identifying and removing seismicity occurring in windows of space and time following (or preceding) a large earthquake. Gardner and Knopoff (1974) define fixed space and time windows that scale with magnitude, whilst Reasenberg (1985) and Musson (1999b) allow the time window to scale according to the foreshock or aftershock activity.

The Reasenberg (1985) algorithm searches for non-Poissonian events in two stages. Following an earthquake not already classified as an aftershock, a spatial interaction zone of

radius r is defined using the following relation:

$$r = rfact * \left(0.4M_W - \left(\frac{\log(\Delta\sigma)}{3} - 1.45\right)\right) \tag{2.13}$$

Where $\Delta\sigma$ is stress drop (assumed to be 30 bars) and rfact is the radius scaling factor (here 10). For all events found to occur within the spatial interaction radius, the expected interval (τ) needed to obtain a P probability of confidence (0.95 in this study) of observing the next event in the aftershock sequence is defined as:

$$\tau = \frac{-\ln(1-P)\,t}{10^{2(\Delta M_W - 1)/3}} \tag{2.14}$$

Where t is the time elapsed since the mainshock and ΔM_W the magnitude difference between the largest earthquakes in the sequence and the completeness magnitude. Since τ is unbounded, an upper bound of 10 days is implemented. The number of aftershocks identified is found to be largely insensitive to the value of stress drop and P. Many of the parameters used here are the same as those of Reasenberg (1985), and can be found in subsequent implementations (Wiemer, 2001; Beauval *et al.*, 2006a). Errors on epicentres and depths are 1.5 km and 2 km respectively. The algorithm to undertake the declustering comes from the ZMAP suite of software (Wiemer, 2001).

A total of 1,285 events are identified as non-Poissonian events within either a foreshock or aftershock sequence, which are then removed from the catalogue. The resulting declustered catalogue contains 8122 earthquakes. The spatial distribution of the full and declustered catalogues can be seen in Figures 2.4 and 2.5 respectively.

2.4 Maximum Magnitude (M_{MAX})

Many of the common magnitude-frequency relations used in seismic hazard analysis suffer from the absence of an upper bound magnitude. In its traditional form, the Gutenberg-Richter relation is one such example. Without an upper bound magnitude (M_{MAX}), these relations allow for a small but finite probability of an unfeasibly large earthquake. In the Monte Carlo method of seismic hazard assessment, where catalogues with effective lengths of millions of years are simulated, this is an important concern. It is therefore necessary to define M_{MAX}. Where knowledge of faulting is extensive, it may be possible to use the physical fault dimensions as a constraint to M_{MAX}. Elsewhere, statistical techniques must be invoked.

Statistical methods for estimating M_{MAX} are useful, but may often produce unphysical results depending on the earthquake catalogue and the parameters used in the calculation. An example of this might be an upper bound magnitude that is lower than the maximum observed magnitude. Alternatively, an ill-fitting distribution may suggest M_{MAX} greater than M_W 11 or 12, which is greater than the finite breaking strength of rock. When

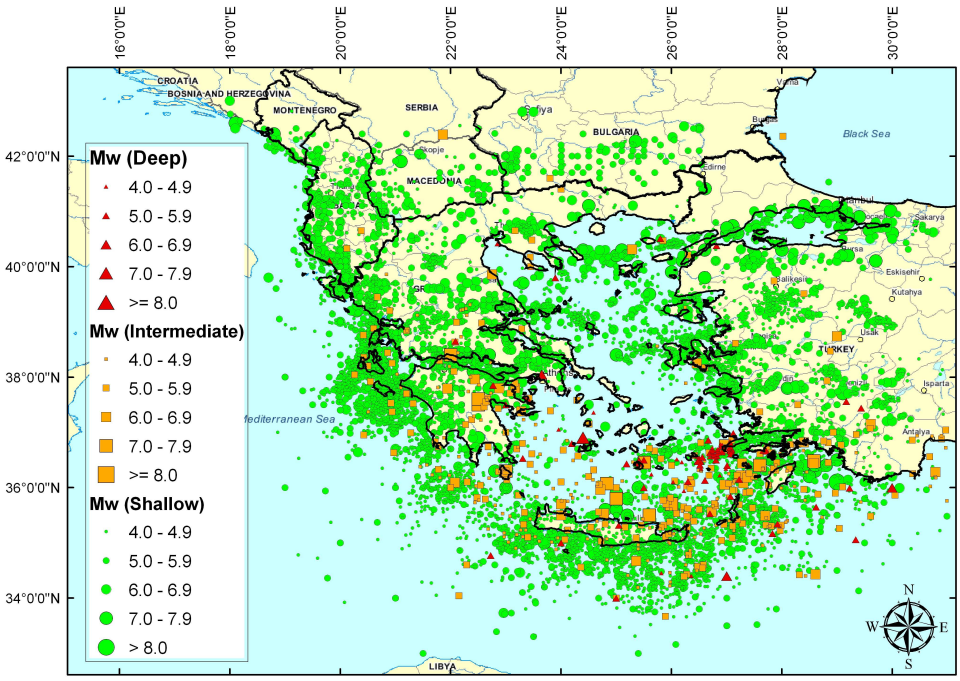

Figure 2.4: Spatial Distribution of the Full Earthquake Catalogue. Green circles indicate crustal events (Depth (km) $\leq$ 60km), yellow circles indicate intermediate depths (60 $<$ Depth (km) $\leq$ 120), and red circles deep earthquakes (Depth (km) $>$ 120)

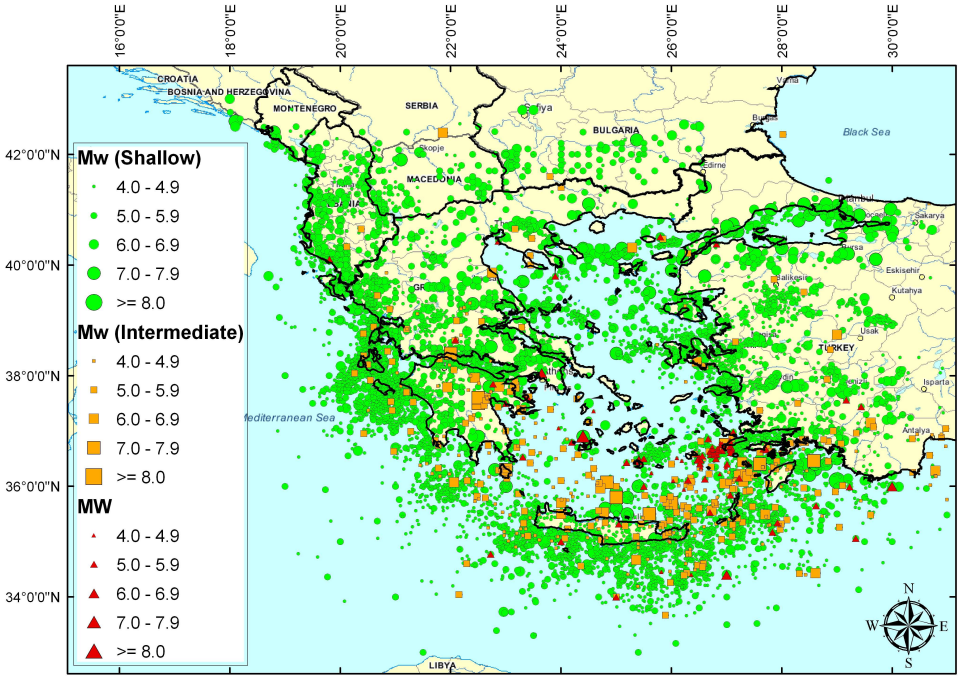

Figure 2.5: Spatial Distribution of the Declustered Earthquake Catalogue. Green circles indicate crustal events (Depth (km) $\leq$ 60km), yellow circles indicate intermediate depths (60 $<$ Depth (km) $\leq$ 120), and red circles deep earthquakes (Depth (km) $>$ 120)

analysing spatial variation in M_{MAX}, either by a grid or on a zone-by-zone basis, algorithmic approaches to delineating M_{MAX} may produce spurious results. The maximum regional magnitude is calculated here using a variety of methods.

2.4.1 Extreme Value Analysis using the Gumbel Type III Distribution

Detailed descriptions of this method and its application to the Aegean region are found in Burton (1979), Makropoulos and Burton (1985a) and Burton *et al.* (2004b), and hence need not be discussed here. Taking annual or multi-year extreme earthquakes, a Gumbel Type III extreme value distribution is fit to the extreme earthquakes via non-linear least squares. The shape of the distribution is:

$$P\left(M_S\right) = G^{III}\left(M_S\right) = \exp\left[-\left(\frac{\omega - M_S}{\omega - \mu}\right)^{1/\lambda}\right] \tag{2.15}$$

The distribution is described by three parameters: ω (the upper bound asymptote), μ (the characteristic or "most likely" annual maximum) and λ (a dimensionless parameter describing the curvature of the distribution). The upper bound asymptote describes the earthquake with a zero probability of being exceeded, i.e. the maximum magnitude. When applied to surface wave magnitudes across the whole Aegean region, assuming Gringorten (1963) plotting point probability and truncated at M_S 5.0 for the period 1901 - 1999, this gives an estimate of M_{MAX} to be approximately equal M_S 8.67 ± 0.2 (Figure 2.6a).

The Gumbel Type III distribution can be analysed spatially by fitting it to discrete squares of earthquakes across a regional grid. This is done in Burton *et al.* (2004b), who fit the Gumbel distribution using a scanning cell of $2^\circ \times 2^\circ$ at a spacing of 0.5°. The spatial distribution of ω (using spline interpolation) can be seen in Figure 2.6b. For much of continental Greece and Turkey, maximum magnitudes fall into the range $8 \leq M_S < 9$. However, some of the peaks in ω are found in low seismicity regions such as the south Aegean and the Greece-Turkey border. These peaks are likely to be due to localised transient features of the earthquake catalogue and may not necessarily be physically accurate. There are also regions where ω is lower than M_S 6. Given the size of large historical earthquakes in low seismicity regions, such as the 1995 Kozani-Grevena event (M_W 6.5), these values of M_{MAX} are likely to be an underestimate.

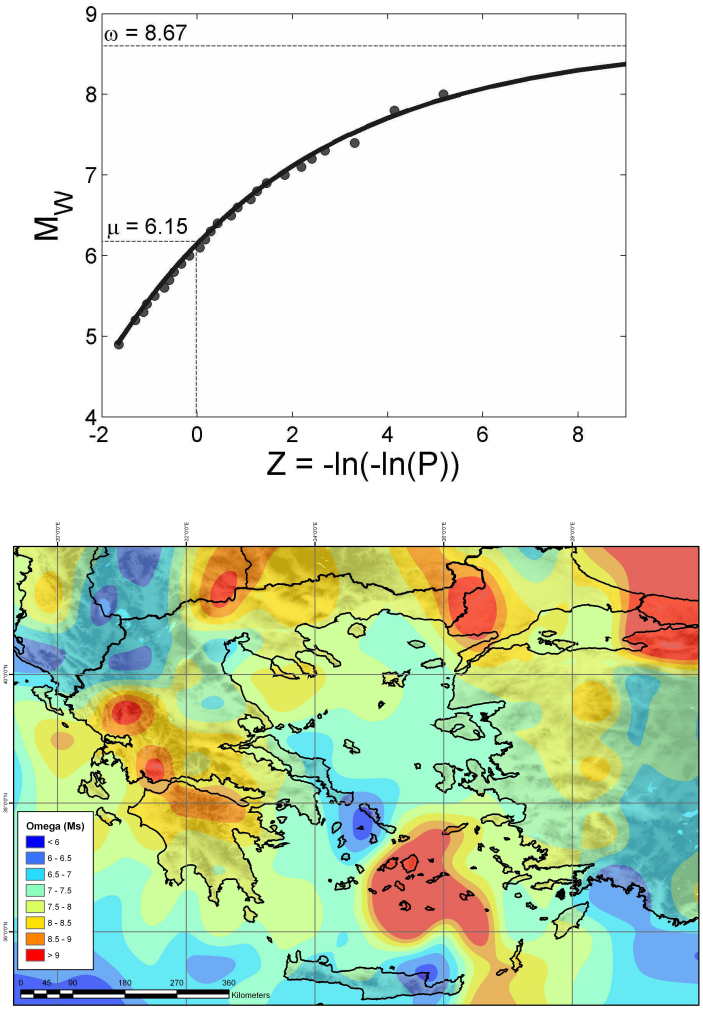

Figure 2.6: a) Gumbel Extreme value plot for the Aegean region 1901 1999, b) Spatial variation in ω of the Gumbel Type III distribution for the Aegean. Data are taken from that printed in Burton *et al.* (2004b) and interpolated using a spline method.

2.4.2 Cumulative Strain Energy

The Gumbel method is clearly useful for analysing spatial variation in return periods of large events, allowing for upper bound magnitude constraints. Whilst the upper bound ω is a useful physical property, it is a statistical parameter and is consequently associated with an uncertainty (σ_ω). The uncertainty on the upper bound can be considerable depending on the data available. Consequently it is capable of returning upper bound magnitudes that may be enormously unphysical if used as a proxy for upper bound earthquake magnitude. It is clearly subject to statistical artefact arising from transient features in the catalogue, which is especially common in low seismicity regions. A more robust method of estimating M_{MAX} is required for the purpose of seismic hazard analysis here. The cumulative strain energy method of Makropoulos and Burton (1983), here using cumulative moment, may provide a more robust estimate for upper bound mag-

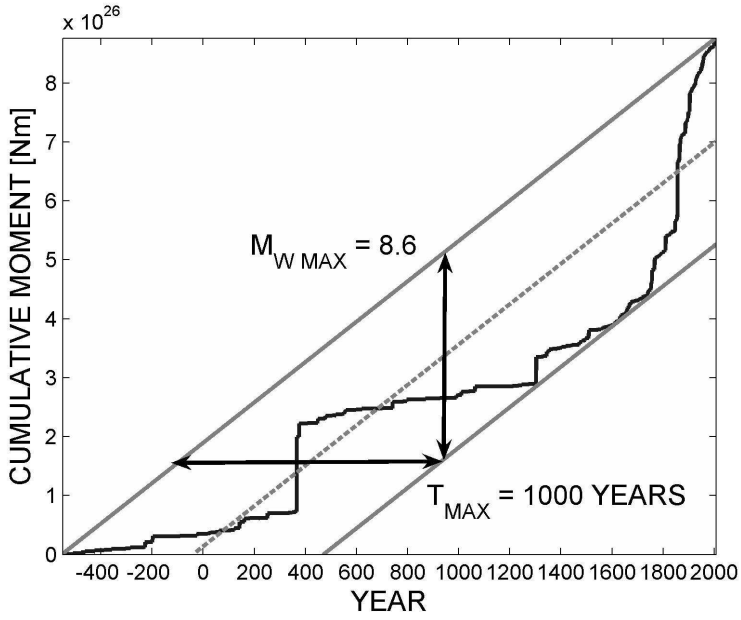

Figure 2.7: Cumulative Moment plot for the Full Aegean catalogue

nitude. The cumulative coseismic moment release from earthquakes over the period 550 B.C. to 1999 A.D. is plotted in Figure 2.7. Two lines, whose gradients represent the mean rate of moment release, are plotted at tangents to the highest and lowest points of the cumulative moment curve. The difference in coseismic moment between these two lines represents an upper bound to the coseismic moment release. This is then converted to M_{MAX} using the formula of Hanks and Kanamori (1979):

$$M_W = \frac{2}{3}\left(\log_{10} M_O - 9.1\right) \tag{2.16}$$

This method has several advantages over many other estimates of M_{MAX}. Firstly it is very simple to implement and computationally efficient. It is essentially non-parametric as it is not dependent on the fit of a probability distribution to the cumulative magnitude-frequency of the catalogue. It also has the advantage that it cannot estimate M_{MAX} lower than the maximum observed magnitude. From a seismological perspective, it is clear from Figure 2.7 that moment release is dominated by the largest earthquakes. This means that it is less sensitive to the choice of cut-off magnitude than other methods. These features make it ideal for applying to seismotectonic zones for the purpose of seismic hazard analysis. It is therefore the method implemented algorithmically in the seismic hazard analysis shown in the following chapters.

The regional maximum observed magnitude for the Aegean is M_W 8.3, and the maximum magnitude estimated using the cumulative moment M_W 8.6. Using M_S and cumulative strain energy, a maximum M_S of 8.6 is obtained, thus confirming the reasonable equivalence suggested in Papazachos and Papazachou (1997). This value is within the likely range suggested as a regional upper bound magnitude when fitting the Gumbel Type III extreme value distribution. Thus there are two separate methods giving a very

similar estimate of M_{MAX}, increasing confidence that it is a reasonable estimation of regional maximum magnitude.

The maximum magnitude determined by this method is, however, dependent on the duration of the catalogue, and the extent to which larger magnitude earthquakes are represented. It is clear from Figure 2.7 that the shape of the cumulative moment plot, and consequently M_{MAX}, is strongly influenced by the occurrence of the M_W 8.3 event in AD 365. This has not taken into account uncertainty regarding the magnitude of this event, which since it is estimated from historical data could be as much as 0.5 magnitude units (Bender, 1983). The historical record of Aegean earthquakes mean that M_{MAX} determined from cumulative moment release may be a reasonable estimate. However, if considering only the 1901 - 2005 period, moment release is effectively dominated by the 1999 Izmit event, giving an estimated regional M_{MAX} equivalent to the maximum observed magnitude.

2.4.3 Alternative Methods of Estimating M_{MAX}

The extreme value and cumulative moment methods implemented previously are just two methods of many that could be used to infer M_{MAX} statistically. These methods both have the advantage that they are not dependent on the fit of a whole process magnitude-frequency relation. Other methods that are dependent on this could also be used. The first is the method of Smith (1976), which utilises both cumulative moment and b-value:

$$M_{MAX} = \frac{\log\left[\left(\frac{1.5}{1.5-b}\right) T\bar{M}_O\right] - 16.0}{1.5} \tag{2.17}$$

where $T\bar{M}_O$ is the cumulative moment of all earthquakes in the time interval T. When applied to the full Aegean catalogue this gives a maximum magnitude of M_W 9.2. Given the error in estimating regional b-value, which itself is dependent on the regional M_C, this high M_{MAX} is likely to be an artefact of the earthquake catalogue. As the cumulative strain energy method of Makropoulos and Burton (1983) does not require calculation of b-value, in algorithmic implementation it is preferred over the Smith (1976) method.

Kijko (2004) suggests several maximum likelihood estimates of M_{MAX}. For an assumed Gutenberg-Richter relation the maximum likelihood method of calculating M_{MAX} is (Kijko and Sellevoll, 1989; Kijko, 2004):

$$M_{MAX} = M_{MAX}^{OBS} + \frac{E_1(n_2) - E_1(n_1)}{\beta \exp(-n_2)} + M_C \exp(-n) \tag{2.18}$$

where $\beta = b\ln(10)$, n is the number of earthquakes above M_C, $n_1 = n/\{1 - \exp[-\beta(M_{MAX} - M_C)]\}$, $n_2 = n_1 \exp[-\beta(M_{MAX} - M_C)]$ and $E_1(z)$ is the ex-

ponential integral function, which is approximated as:

$$E_1(z) \approx \frac{z^2 + 2.334733z + 0.250621}{z(z^2 + 3.330657z + 1.681534)} \exp(-z) \tag{2.19}$$

When applied to the Aegean catalogue (1901 - 2005), with M_C 5.2, this gives an M_{MAX} of M_W 8.6$\pm$0.1. This is, again, in good agreement with the cumulative moment and Gumbel methods of estimating M_{MAX}. This method is clearly dependent on b-value, but adjusts for the completeness magnitude. Although iterative, it is computationally efficient and tends to converge in fewer than 100 iterations.

Equation 2.18 is derived assuming a Gutenberg-Richter relation with fixed b-value. When using the whole regional catalogue there are a sufficient number of earthquakes to constrain b-value with reasonable accuracy. When analysing M_{MAX} for smaller catalogues, the errors in b-value then propagate into the M_{MAX} calculation. This means that it is possible for this method to give M_{MAX} less than the maximum observed magnitude. Kijko (2004) offers two other likelihood methods, one allowing for uncertainty in b-value and one non-parametric Gaussian method. Both methods drastically increase computational expense, and in the case of the non-parametric Gaussian method, introduce a kernal smoothing factor that is itself poorly constrained and must be solved by numerical integration. Although these methods may be adequate when considering a single determination of M_{MAX} for a regional catalogue, they are unsuitable for spatial analysis of M_{MAX}.

The cumulative moment method of estimating maximum magnitude (Makropoulos and Burton, 1983) is preferred, in this study at least, largely because it is not explicitly dependent on b-value. This is a useful property as it is not affected by the uncertainty that arises in b-value for many catalogues. This can produce some erroneous results when considering spatial variation in M_{MAX}, Other statistical methods (Smith, 1976; Kijko and Sellevoll, 1989) do have an explicit dependence on b-value. Despite the different assumptions in each method, three of the four methods considered here give a regional M_{MAX} of approximately M_W 8.6. The good agreement between the methods using both whole process and partial process statistics suggest this estimate of M_{MAX} can be treated with considerable confidence.

2.5 Analysis of the Gutenberg-Richter Relation fit for the Aegean Catalogue

2.5.1 Definition of Subsets

The initial analysis of catalogue completeness using the Stepp (1971) method clearly shows how completeness magnitude decreases as the catalogue becomes more recent.

To determine a- and b-value from a single catalogue with time-variable completeness is a difficult task, and ultimately may misrepresent the real seismic behaviour of earthquakes in the study region. Instead, the Aegean catalogue is divided into six subsets (seven if the entire catalogue is counted as a subset):

1. SUBSET 1 (Historical): All earthquakes in the period 550 B.C. to 1899 A.D.
2. SUBSET 2 (Early Instrumental): Earthquakes in the period 1900 - 1963 A.D.
3. SUBSET 3 (ISC): Earthquakes in the period 1964 - 1998 A.D.
4. SUBSET 4 (NOA): Earthquakes in the period 1999 - 2005 A.D.
5. SUBSET 5 (All Instrumental): Earthquakes in the period 1900 - 2005 A.D.
6. SUBSET 6 (Modern Instrumental): Earthquakes in the period 1964 - 2005 A.D.
7. SUBSET 7 (Entire): All earthquakes in the period 550 B.C. to 2005 A.D.

For each of these subsets completeness magnitude is determined using three methods of completeness discussed in sections 2.2.2 - 2.2.4.

2.5.2 Fit of the Gutenberg-Richter parameters (a- and b-value)

There are a variety of methods to determine a- and b-value that are used. Generally these fall into two categories: least squares and maximum likelihood. Each of the two methods will, for most real data sets, produce different estimates of a- and b-value. In the simple linear least squares method the value of $\log\ N_C$ for each magnitude is treated equally (also assuming independent normal variance for each magnitude). This will tend to produce fits to the data that appear graphically pleasing as they are essentially lines of best fit. However, the least squares method will be more strongly influenced by the larger more variable magnitudes, which will depend strongly on the duration of the catalogue (Giardini *et al.*, 2004). This can be compensated for by implementing a weighted least squares approach. In the absence of an appropriate weighting scheme, however, the maximum likelihood method is often preferred.

The maximum likelihood method (Aki, 1965; Utsu, 1965; Page, 1968) calculates b-value from the following relation, assuming magnitude bins of 0.1 magnitude units:

$$b = \frac{\log_{10}(e)}{(\bar{m} - M_c + 0.05)} \tag{2.20}$$

where $\bar{m}$ is the mean magnitude, M_C the threshold magnitude (here the completeness magnitude) and e is the natural exponent. Uncertainty on b-value is determined from

the Shi and Bolt (1982) formula:

$$\sigma_b = 2.3b^2\sqrt{\frac{\sum_{i=1}^{N}(m_i - \bar{m})^2}{N(N-1)}} \tag{2.21}$$

N is the total number of earthquakes greater than or equal to M_C. As b-value is dependent on the mean magnitude, then the large number of small events will have a greater influence. This makes the b-value less sensitive to large earthquakes and consequently more robust. In doing so, however, it may not appear to fit the distribution as well as the least squares estimates, and is more strongly influenced by completeness magnitude than the least squares estimates of b-value.

The parameters M_C, a- and b-value and their respective uncertainties (for both least squares and maximum likelihood methods) are shown in Tables 2.2 and 2.3, and the cumulative number of events verses M_W plotted in Figures 2.8 and 2.9. Both the full and declustered catalogues are compared. It should be quite apparent that M_C estimated from each of the different methods can differ to a great extent depending on the catalogue at hand.

Many data sets display obvious disparity between the b-values estimated by least squares and maximum likelihood methods. For most of the subsets the maximum likelihood b-value is lower, which tends to result in overestimation of the number of large earthquakes. The only consistent exception to this is SUBSET 4, which consists of earthquakes in the period 1999 - 2005 A.D. The reason for this exception is the inclusion of the 1999 Izmit earthquake, which has a significant influence on the least squares estimate but not on the maximum likelihood estimate. As each of the two methods of b-value estimation is influenced by a different part of the catalogue, it is not unreasonable to suggest that a good indication of the "true" b-value may be found where both methods are in good agreement. Here, the best agreement between methods can be found in SUBSET 5 (1900 - 2005), when $M_C \approx 5.2$. This particular subset is of interest as it spans most of the "instrumental period". The b-value found for this period, using the full catalogue, is 1.31 ± 0.03 (Least Squares) and 1.33 ± 0.03 (MLE), with a-values of 8.14 ± 0.205 and 8.43 ± 0.14, respectively. For the declustered catalogue, b-values for the same subset are 1.264 ± 0.024 (Least Squares) and 1.206 ± 0.026 (MLE), with a-values of 7.878 ± 0.156 and 7.539 ± 0.137, respectively.

Table 2.2: Gutenberg-Richter Parameters for the listed subsets of the full Aegean catalogue

						Least Squares				Maximum Likelihood			
	START	FINISH	N	M_C Type	M_C	a	σa	b	σb	a	σa	b	σb
SUBSET1	-550	1899	419	Max. Curve.	7	8.727	0.496	1.462	0.065	11.014	1.384	1.774	0.198
				GOF90	6.8	8.671	0.372	1.454	0.049	7.974	0.68	1.354	0.1
				b-stability	7	8.727	0.496	1.462	0.065	11.014	1.384	1.774	0.198
SUBSET2	1900	1963	1971	Max. Curve.	5.5	8.41	0.416	1.339	0.063	10.437	0.441	1.698	0.08
				GOF90	5.7	8.458	0.509	1.346	0.076	6.699	0.356	1.065	0.062
				b-stability	5.7	8.458	0.509	1.346	0.076	6.699	0.356	1.065	0.062
SUBSET3	1964	1998	2158	Max. Curve.	4.8	7.977	0.315	1.314	0.054	7.209	0.146	1.171	0.03
				GOF90	4.8	7.977	0.315	1.314	0.054	7.209	0.146	1.171	0.03
				b-stability	5.5	8.663	0.689	1.422	0.111	6.49	0.383	1.057	0.07
SUBSET4	1999	2005	4356	Max. Curve.	4	6.878	0.264	1.108	0.045	8.784	0.092	1.497	0.023
				GOF90	4	6.878	0.264	1.108	0.045	8.784	0.092	1.497	0.023
				b-stability	4	6.878	0.264	1.108	0.045	8.784	0.092	1.497	0.023
SUBSET5	1900	2005	8485	Max. Curve.	4	6.627	0.201	1.078	0.034	4.305	0.021	0.599	0.005
				GOF90	5.2	8.141	0.205	1.308	0.032	8.426	0.142	1.329	0.027
				b-stability	5.7	8.157	0.331	1.31	0.05	6.915	0.302	1.111	0.053
SUBSET6	1964	2005	6514	Max. Curve.	4	7.216	0.159	1.197	0.027	5.677	0.04	0.872	0.01
				GOF90	4	7.216	0.159	1.197	0.027	5.677	0.04	0.872	0.01
				b-stability	4	7.216	0.159	1.197	0.027	5.677	0.04	0.872	0.01
SUBSET7	-550	2005	8904	Max. Curve.	4	4.498	0.205	0.888	0.033	2.671	0.02	0.532	0.005
				GOF90	5	5.641	0.254	1.05	0.038	3.763	0.057	0.73	0.011
				b-stability	4.1	4.59	0.21	0.902	0.033	2.595	0.02	0.517	0.005

Table 2.3: Gutenberg-Richter Parameters for the listed subsets of the declustered Aegean catalogue

						Least Squares				Maximum Likelihood			
	START	FINISH	N	M_C Type	M_C	a	σa	b	σb	a	σa	b	σb
SUBSET1	-550	1899	418	Max. Curve.	7	8.744	0.494	1.464	0.065	10.962	1.369	1.766	0.196
				GOF90	6.8	8.68	0.37	1.456	0.049	7.954	0.676	1.351	0.099
				b-stability	7	8.744	0.494	1.464	0.065	10.962	1.369	1.766	0.196
SUBSET2	1900	1963	1565	Max. Curve.	5.5	7.937	0.285	1.26	0.043	9.58	0.401	1.55	0.073
				GOF90	5.6	7.889	0.313	1.253	0.047	7.596	0.364	1.212	0.065
				b-stability	5.3	7.94	0.237	1.261	0.036	8.024	0.203	1.278	0.038
SUBSET3	1964	1998	1861	Max. Curve.	4.8	7.577	0.324	1.244	0.055	6.887	0.15	1.116	0.031
				GOF90	4.8	7.577	0.324	1.244	0.055	6.887	0.15	1.116	0.031
				b-stability	4.8	7.577	0.324	1.244	0.055	6.887	0.15	1.116	0.031
SUBSET4	1999	2005	3769	Max. Curve.	4	6.748	0.232	1.083	0.039	8.519	0.096	1.447	0.024
				GOF90	4	6.748	0.232	1.083	0.039	8.519	0.096	1.447	0.024
				b-stability	4	6.748	0.232	1.083	0.039	8.519	0.096	1.447	0.024
SUBSET5	1900	2005	7195	Max. Curve.	4	6.4	0.186	1.042	0.031	4.223	0.024	0.597	0.006
				GOF90	5.2	7.878	0.156	1.264	0.024	7.539	0.137	1.206	0.026
				b-stability	5.3	7.904	0.169	1.268	0.026	7.938	0.176	1.277	0.033
SUBSET6	1964	2005	5630	Max. Curve.	4	7.037	0.167	1.167	0.028	5.538	0.042	0.853	0.01
				GOF90	4	7.037	0.167	1.167	0.028	5.538	0.042	0.853	0.01
				b-stability	4	7.037	0.167	1.167	0.028	5.538	0.042	0.853	0.01
SUBSET7	-550	2005	7613	Max. Curve.	4	4.335	0.212	0.865	0.034	2.557	0.021	0.521	0.005
				GOF90	5	5.521	0.264	1.033	0.039	3.377	0.055	0.668	0.011
				b-stability	4.1	4.427	0.218	0.878	0.035	2.478	0.022	0.505	0.005

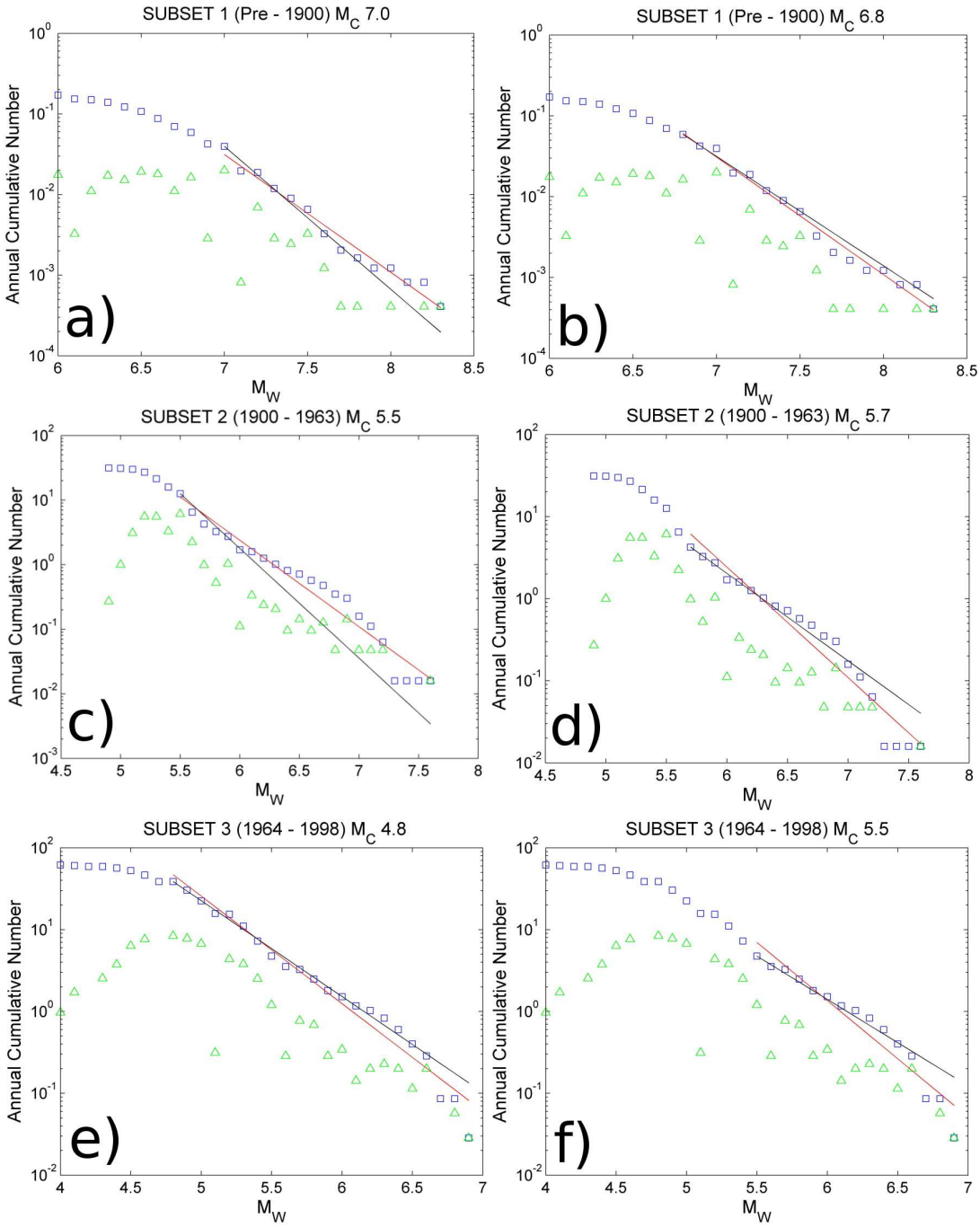
SUBSET 1 (Pre - 1900) M_C 7.0
SUBSET 1 (Pre - 1900) M_C 6.8
SUBSET 2 (1900 - 1963) M_C 5.5
SUBSET 2 (1900 - 1963) M_C 5.7
SUBSET 3 (1964 - 1998) M_C 4.8
SUBSET 3 (1964 - 1998) M_C 5.5
Annual Cumulative Number
M_W
a)
b)
c)
d)
e)
f)

Figure 2.8

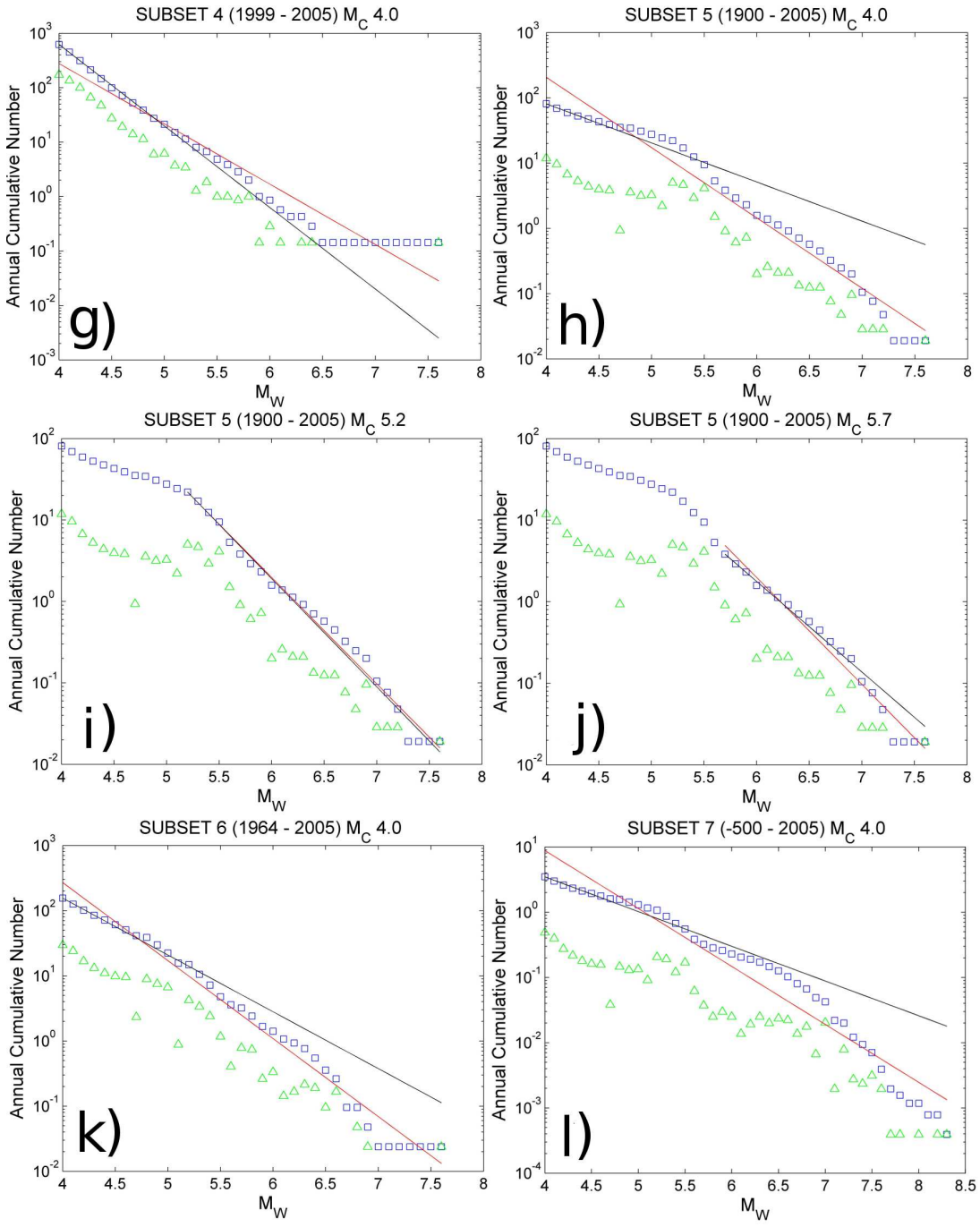
SUBSET 4 (1999 - 2005) M_C 4.0
SUBSET 5 (1900 - 2005) M_C 4.0
SUBSET 5 (1900 - 2005) M_C 5.2
SUBSET 5 (1900 - 2005) M_C 5.7
SUBSET 6 (1964 - 2005) M_C 4.0
SUBSET 7 (-500 - 2005) M_C 4.0
Annual Cumulative Number
M_W
g)
h)
i)
j)
k)
l)

Figure 2.8

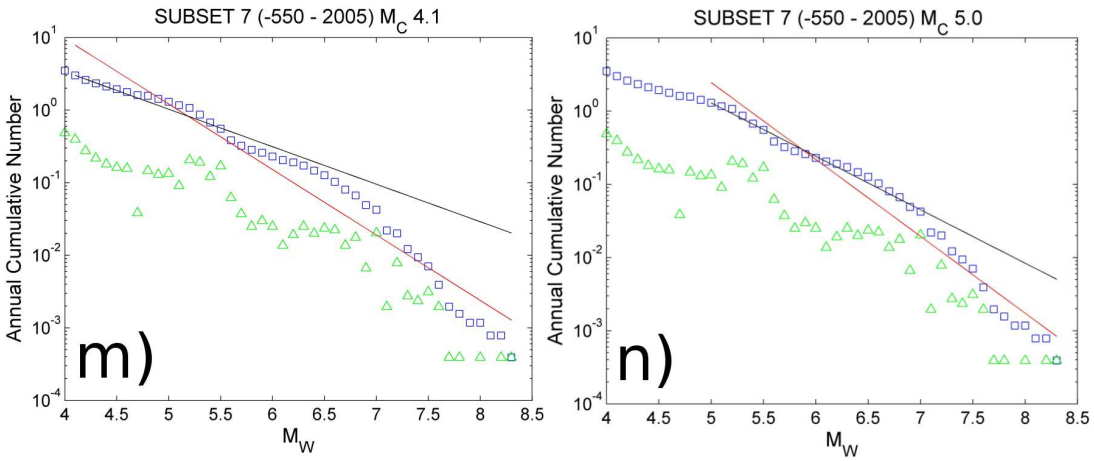

Figure 2.8: Plots of absolute (green triangle) and cumulative (blue square) annual numbers of earthquakes in the full Aegean catalogue verses M_W, with Gutenberg-Richter relation fits given. Red lines indicate least squares fit, whilst black lines represent the maximum likelihood fit. a) Subset 1 (M_C 7 - Max. Curve & b-stability), b) Subset 1 (M_C 6.8 - GOF90), c) Subset 2 (M_C 5.5 - Max. Curve), d) Subset 2 (M_C 5.7 - GOF90 & b-stability), e) Subset 3 (M_C 4.8 - Max. Curve & GOF90), f) Subset 3 (M_C 5.5 - b-stability), g) Subset 4 (M_C 4 - all methods), h) Subset 5 (M_C 4 - Max. Curve.), i) Subset 5 (M_C 5.2 - GOF90), j) Subset 5 (M_C 5.7 - b-stability), k) Subset 6 (M_C 4.0 - all methods), l) Subset 7 (M_C 4 - Max. Curve.), m) Subset 7 (M_C 5 - GOF90) and n) Subset 7 (M_C 4.1 - b-stability)

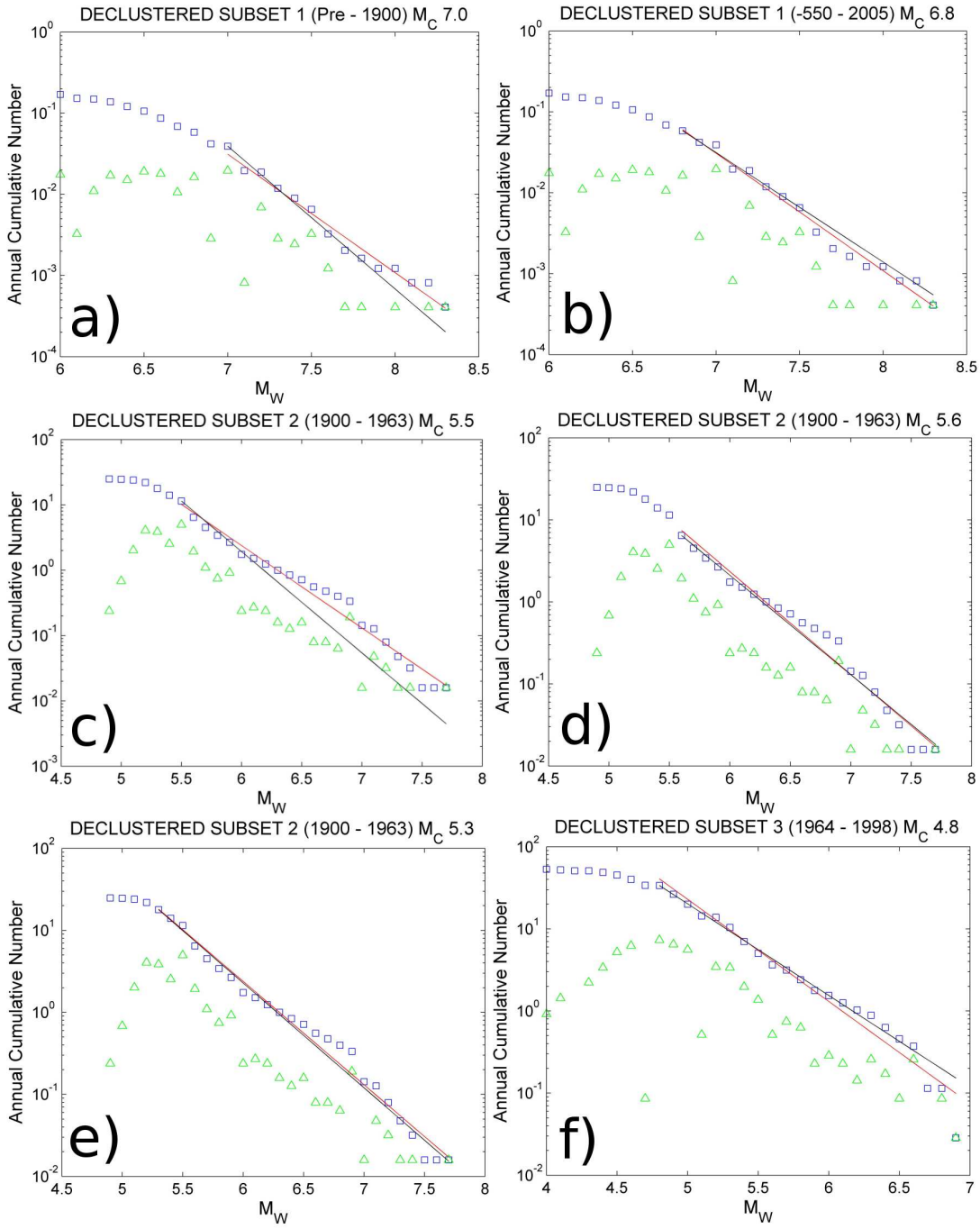
DECLUSTERED SUBSET 1 (Pre - 1900) M_C 7.0
DECLUSTERED SUBSET 1 (-550 - 2005) M_C 6.8
DECLUSTERED SUBSET 2 (1900 - 1963) M_C 5.5
DECLUSTERED SUBSET 2 (1900 - 1963) M_C 5.6
DECLUSTERED SUBSET 2 (1900 - 1963) M_C 5.3
DECLUSTERED SUBSET 3 (1964 - 1998) M_C 4.8
Annual Cumulative Number
M_W
a)
b)
c)
d)
e)
f)

Figure 2.9

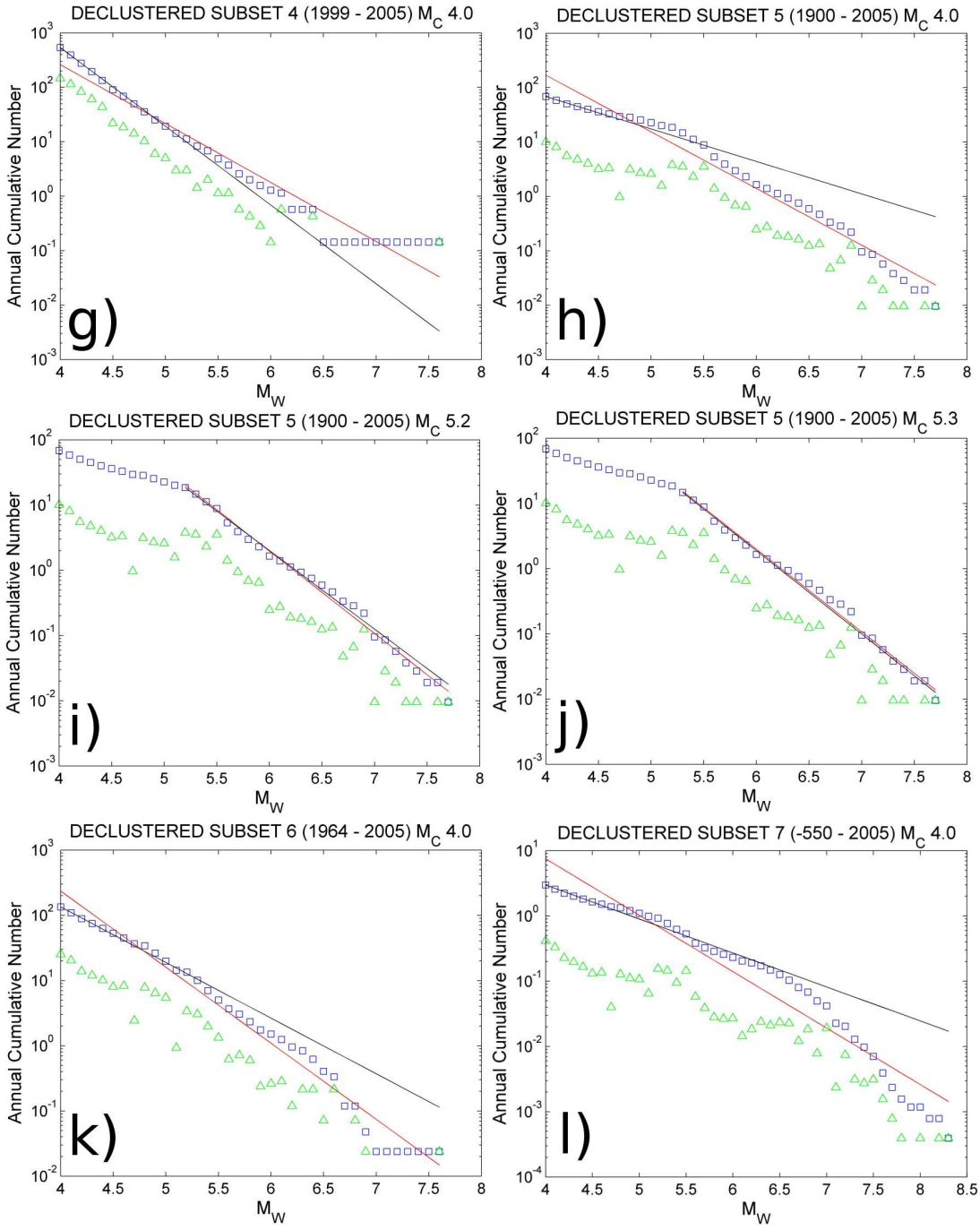
DECLUSTERED SUBSET 4 (1999 - 2005) M_C 4.0
DECLUSTERED SUBSET 5 (1900 - 2005) M_C 4.0
DECLUSTERED SUBSET 5 (1900 - 2005) M_C 5.2
DECLUSTERED SUBSET 5 (1900 - 2005) M_C 5.3
DECLUSTERED SUBSET 6 (1964 - 2005) M_C 4.0
DECLUSTERED SUBSET 7 (-550 - 2005) M_C 4.0
Annual Cumulative Number
M_W
g)
h)
i)
j)
k)
l)

Figure 2.9

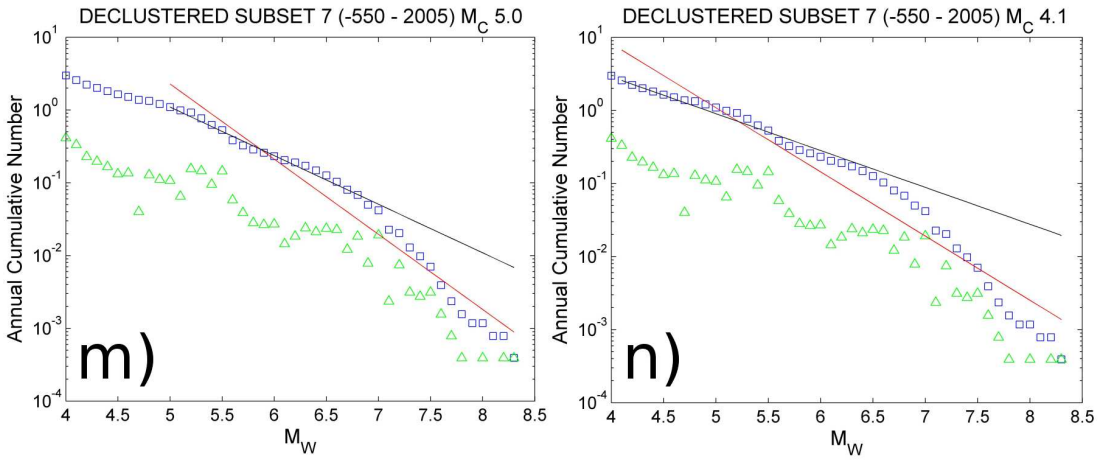

Figure 2.9: Plots of absolute (green triangle) and cumulative (blue square) annual numbers of earthquakes in the declustered Aegean catalogue verses M_W, with Gutenberg-Richter relation fits given. Red lines indicate least squares fit, whilst black lines represent the maximum likelihood fit. a) Subset 1 (M_C 7 - Max. Curve & b-stability), b) Subset 1 (M_C 6.8 - GOF90), c) Subset 2 (M_C 5.5 - Max. Curve), d) Subset 2 (M_C 5.6 - GOF90), e) Subset 2 (M_C 5.3 - b-stability), f) Subset 3 (M_C 4.8 - all methods), g) Subset 4 (M_C 4 - all methods), h) Subset 5 (M_C 4 - Max. Curve.), i) Subset 5 (M_C 5.2 - GOF90), j) Subset 5 (M_C 5.3 - b-stability), k) Subset 6 (M_C 4.0 - all methods), l) Subset 7 (M_C 4 - Max. Curve.), m) Subset 7 (M_C 5 - GOF90) and n) Subset 7 (M_C 4.1 - b-stability)

For the regional catalogue these b-values appear to be towards the higher end of the range expected from other regional analyses of b-values (Frohlich and Davis, 1993). It is not entirely clear whether this is due to the seismotectonics of the region, or an artefact of the catalogues in question. What is noticeable, however, is that for the complete section of the three subsets of earthquakes post-1900 A.D., the Gutenberg-Richter parameters are very consistent. It is the historical subset (pre - 1900) that differs significantly in terms of both a- and b-value. Figure 2.10 shows trend lines described by the equations:

$$\log_{10}(N_c) = 7.7 - 1.25M_W \tag{2.22}$$

and

$$\log_{10}(N_c) = 7.5 - 1.20M_W \tag{2.23}$$

for the full and declustered versions respectively. These lines describe the post-1900 distribution of earthquakes well. The historical subset, however, lies below this line, producing a b-value of 1.5 and a-value of approximately 9.12. Note that the Reasenberg (1985) algorithm of declustering removes only one earthquake from this period, hence the differences in a- and b-value between the full and declustered catalogue are trivial.

The b-value of the historical period is higher than that of the 20th century. There are many plausible explanations for this observation. As errors in the magnitude estimates for the largest earthquakes have not been taken into account, it is possible that the very largest events ($M_W > 8.0$) were underestimated. There is insufficient information, however,

to determine whether this is the case. Assuming it is not, Figure 2.10 indicates that the annual number of small earthquakes is lower than that expected from the behaviour of larger earthquakes. Alternatively, it could be perceived that the annual number of large earthquakes is also less than that expected from the behaviour of smaller earthquakes.

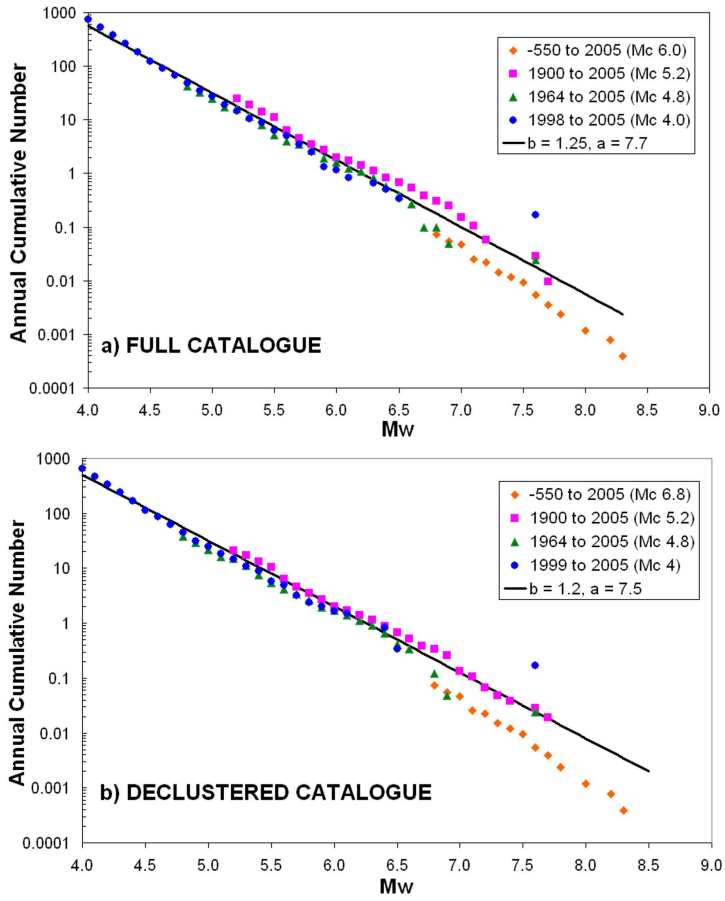

Figure 2.10: Combined G-R plots for separate durations of the a) Full and b) Declustered catalogue at different levels of completeness. 550BC to 2005 AD (M_C 6.8 - Orange Diamonds), 1900 AD to 2005 AD (M_C 5.2 Pink Squares), 1964 AD to 2005 AD (M_C 4.8 Green Triangles) and 1999 AD to 2005 AD (M_C 4.0 blue circles)

There also remains the possibility that Aegean seismicity displays time-dependent behaviour. This would imply that the number of large earthquakes observed in the 20th century is greater than that expected from analysis of the last two millennia. Evidence of time-dependence in Greek seismicity from Rescaled Range analysis (Xu and Burton, 2001, 2006) suggests a process with a long-memory, but does not reject the occurrence of this process by chance.

2.5.3 Spatial Variation in b-value

Two plots of spatial variation in b-value are shown in 2.11 . In both of these plots a fixed M_C of M_W 5.2 is used, with plot a) gridded at $0.5^\circ \times 0.5^\circ$ resolution, and plot b) at $0.2^\circ \times 0.2^\circ$ resolution. The general trend of decreasing b-value towards the northeast is clearly

visible. This is consistent with the analyses of Hatzidimitriou *et al.* (1985) and Papazachos (1990). The Hellenic arc displays higher b-values, especially in the region immediately south of Greece. It is interesting that this localised area of high b-value occurs at the north-western end of an area of anomalously low seismicity, broadly corresponding to the rupture of the A.D. 364 earthquake (M_W 8.3). Other localised areas of anomalously high b-value can be seen at the western end of the North Aegean trough, and in the Izmir region of western Turkey.

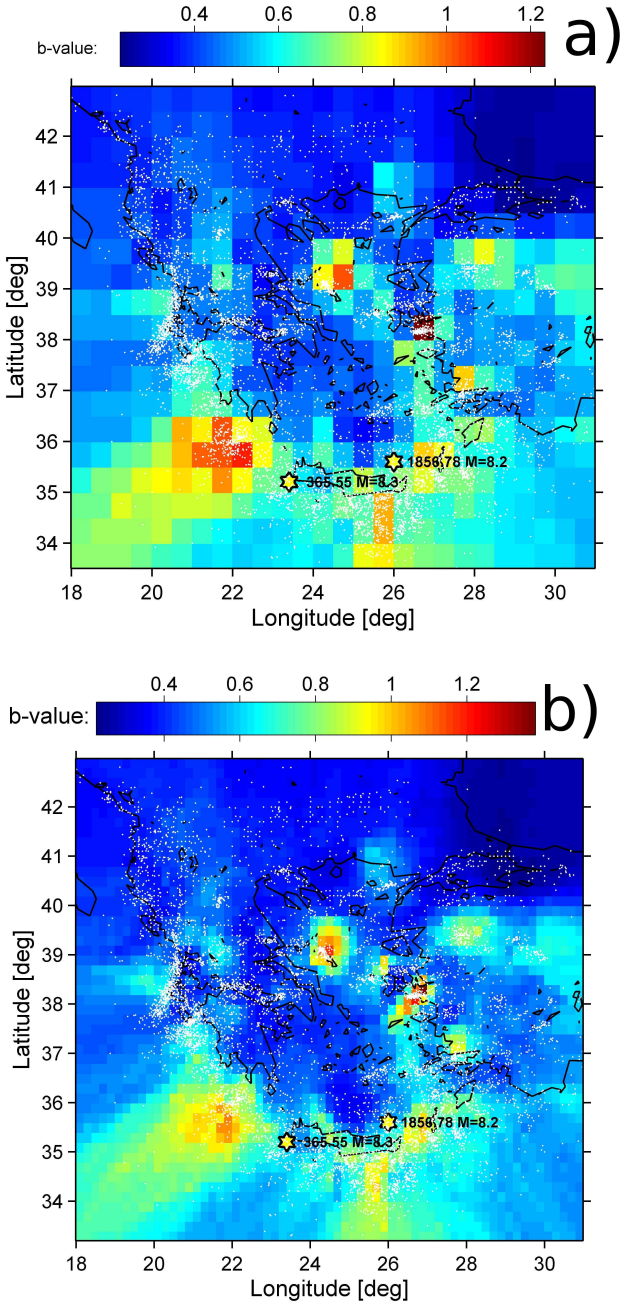

Figure 2.11: Spatial variation in b-value using the full Aegean catalogue 1900 - 2005 with fixed M_C 5.2 plotted at a) $0.5^\circ \times 0.5^\circ$ and b) $0.2^\circ \times 0.2^\circ$ resolution.

It would appear that there is good correspondence between b-value and regions of high deformation. There are some areas where this is not necessarily the case, however. In par-

ticular the Adriatic coast and the Gulf of Corinth are areas of high seismicity that do not display high b-values in these maps. The suggestion is made that higher b-values arise along the Hellenic arc (Hatzidimitriou *et al.*, 1985) due to greater geological heterogeneity (Mogi, 1962) at the plate margin. Conversely lower b-values are found in the older, more structurally stable crust in the north-eastern Aegean region.

The heterogeneity argument may explain the regional trend in b-values but does not provide a full explanation for the spatial b-value distribution. It should also be apparent that some of the features of the b-value distribution may be statistical artefact. The cut-off magnitude of M_W 5.2 means that for many of the lower seismicity regions there are too few events to constrain b-value accurately, which, when using the maximum likelihood method, may provide a bias toward lower b-values. Nevertheless, the high b-values in the north Aegean Sea and along the western Hellenic arc may very well be attributed to physical rather than statistical causes.

2.6 Estimation of Seismicity Rate in Monte Carlo Seismic Hazard Analysis

Although the methods of estimating b-value are well known and quite robust, estimation of a-value is more complex than often assumed. For the purposes of simulation of synthetic catalogues the seismicity rate has an impact not only on hazard but also on the computational performance. The exponential nature of the Gutenberg-Richter relation is such that a difference of 0.1 in regional a-value translates to a difference of tens of thousands of earthquakes for 500 years of synthetic earthquakes for $M_W \geq 4.0$.

When utilising a source zone model, it is necessary to determine seismicity rate, b-value and M_{MAX} for each zone. The main problem is that in trying to ascertain recurrence parameters of a smaller subset of earthquakes, the parameters will have greater errors. Equally, it is likely that the resulting recurrence relation will reflect the brevity of the catalogue in the sense that too few large events may be recorded, or too many. It is common practice in Monte Carlo simulations to use a-value as the basis for the number of earthquakes simulated in each zone (Musson, 1999b; Giardini *et al.*, 2004). If the uncertainty in a-value arising from the observed earthquake catalogue is taken into consideration then the variation in the number of earthquakes in the synthetic catalogue will be very large. The alternative would be to assume a general rate of seismicity for the entire catalogue, based upon the regional recurrence relation. In this case, the number of earthquakes simulated in each zone would be proportioned according to the relative number of earthquakes in each zone.

The approaches to simulating seismicity rate described are reliant on different assumptions. In simulating seismicity on a zone by zone basis, it is assumed that the a-value determined for the zone is true for the entire time period simulated; it is a true reflection

of the rate of stationary seismicity. The likelihood of this assumption depends on the timescale of the complete catalogue and whether this is a true reflection of the seismicity of the zone. The alternative approach of assuming a fixed regional a-value and assigning earthquakes to zones in proportion to the observed distribution makes a different assumption. This method assumes that the observed proportion of earthquakes in each zone reflects the overall spatial variation in seismicity across the region.

An alternative to using a-value is to characterise seismicity rate using observed recurrence times. This is similar to the approach of Ebel and Kafka (1999) who use an incremental time counter to characterise the time distribution of the synthetic catalogues. Assuming that the input catalogue is complete above the minimum magnitude and that non-Poissonian events have been removed, the time counter can be simulated in different ways. The first method makes no assumption about the inter-event time distribution and will generate the time between events by randomly re-sampling, with replacement, the observed inter-event time. This approach will be implemented here. Alternatively, if one wishes to assign a probability distribution to the inter-event time the most obvious candidate is to sample randomly from a Poisson distribution. However, there is considerable evidence to suggest that inter-event time may be adequately represented by a lognormal distribution (Musson *et al.*, 2002) or Brownian Passage time distribution (Matthews *et al.*, 2002). Should one wish to implement an incremental time counter when $M_{MIN} < M_C$ in the seismic hazard analysis, and a-value is the sole indication to seismic rate, the re-sampling approach is insufficient. In such circumstances it may be most prudent to use the Poisson distribution as a means of simulating inter-event time.

2.7 Alternative Recurrence Relations

2.7.1 The Form of the Double-Truncated Gutenberg-Richter Relation

So far, only the Gutenberg-Richter relation and its truncated version have been considered as suitable recurrence relations for use in this analysis. Since it is a necessity in seismic hazard analysis to define upper and lower bounds on magnitude (M_{MAX} and M_{MIN} respectively) the truncated version is used in all subsequent algorithms. The probability density function and cumulative density function of the truncated Gutenberg-Relation are:

$$f(m) = \frac{\beta \exp\left[-\beta\left(m - m_{\min}\right)\right]}{1 - \exp\left[-\beta\left(m_{\max} - m_{\min}\right)\right]} \tag{2.24}$$

where $\beta = b \ln(10)$ and

$$F(m) = \frac{1 - \exp\left[-\beta\left(m - m_{\min}\right)\right]}{1 - \exp\left[-\beta\left(m_{\max} - m_{\min}\right)\right]} \tag{2.25}$$

It has been recognised that earthquake recurrence varies substantially across the globe. The Gutenberg-Richter relation does not always adequately describe such variation. Furthermore it is often unclear to what extent variation in b-value is due to seismotectonic properties or to shortcomings of the catalogue used (Frohlich and Davis, 1993), or the method of calculation. Utsu (1999) notes that where two different regions may have similar b-values the patterns of the frequency-magnitude distribution may still differ. The differences may be elucidated in an alternative distribution function.

A description and comparison of many existing recurrence relations is given in Utsu (1999), who uses the Akaike Information Criterion to assess the performance of these relations for regions of Japan. For most of the regions considered, the Gutenberg-Richter relation and its truncated version still provide the best fit to the data. Such an observation is borne out by Giardini *et al.* (2004), who find no statistical reason to assume any alternative relation to the Gutenberg-Richter relations.

2.7.2 The Characteristic Earthquake Distribution

Before rejecting all other recurrence relations, there is still an important one that must be considered: the "Characteristic earthquake" distribution (Schwartz and Coppersmith, 1984; Youngs and Coppersmith, 1985). This relation assumes an exponential distribution for magnitudes in the range $M_{MIN} \leq M_W < M_{CH}$ and a uniform distribution in the range $M_{CH} \leq M_W \leq M_{MAX}$. This means that for a particular fault, large earthquakes of relatively similar magnitude occur more frequently than is expected from the exponential distribution of smaller magnitudes. This behaviour is often visible in plots of cumulative numbers of earthquakes against magnitude. Many physical interpretations for this behaviour have been suggested (Wesnousky, 1994; Lopez-Ruiz *et al.*, 2004); however, the model is also the subject of considerable controversy. Though the details of the controversy are beyond the scope of this research, there is a strong body of evidence to suggest that the "characteristic earthquake" is largely a statistical artefact (Kagan, 1993; Stein and Newman, 2004; Jackson and Kagan, 2006). A principal criticism of the characteristic earthquake hypothesis lies in the assumption of a segmented fault model. In such a model an active fault is divided into segments, with the characteristic earthquake representing a rupture through the entire segment. Consequently, estimates of the characteristic earthquake magnitude are often derived from physical parameters of the segment, such as fault length, moment rate etc.

Of course one of the greatest sources of consternation in the "characteristic earthquake" model relates to the scale of the source concerned. It is recognised that as the area considered increases, the characteristic earthquake becomes less prominent, eventually disappearing. Conversely, as the source more closely approximates a fault segment, as opposed to a zone, the characteristic earthquake is of course more prominent. Proponents of the characteristic earthquake model would generally argue that this supports the hy-

pothesis and clearly relates the statistical phenomenon to the physics of the fault. Usually this is in the form of moment rate or characteristic segment length. It could also be argued, however, that by refining the earthquake catalogue to a smaller area more transient properties emerge. This suggests that the upper magnitude range of the recurrence relation is dominated by one or two small events, which could also be explained by Poissonian behaviour if considered over a short timescale. For many regions of the world where historical records are short (e.g. California), the likelihood of the record adequately capturing the full cycle of seismicity is small. Incorporation of paleoseismic geological information may be misleading as it is the largest earthquakes that can be most accurately determined. Earthquakes in the intermediate to large range, especially those arising due to static and dynamic stress changes in the years preceding and following a large event, may be indistinguishable from the largest shock. Consequently, earthquakes in this range may be under represented in the catalogue, which will increase the gradient of the Gutenberg-Richter relation thus resulting in a bimodal model.

The characteristic earthquake debate is indeed a lively one, not least of all because of the implications it has regarding the predictability of earthquakes. One suspects that controversy will continue to rage well into the future as earthquake activity across the globe continues both to support and contradict this assertion. The application of the "characteristic earthquake" model to Aegean seismic hazard analysis must be considered, even if it is not ultimately implemented.

2.7.3 Application of Characteristic Earthquake Distribution in the Aegean

The application of alternative recurrence relations to the Aegean region has been undertaken in the past (Main and Burton, 1984, 1989; Papadopoulos *et al.*, 1993, 2003). The study of Main and Burton (1984) develops a Boltzmann distribution of the form:

$$n\left(m\right)dm = C\exp\left[-\beta m - 1.5M_o\left(m\right)\right]dm \tag{2.26}$$

This particular model doesn't incorporate characteristic earthquake behaviour, but is designed to allow for a non-linear decrease in n(m) at higher magnitudes without a sudden cut-off. It is the application of the Lomnitz-Adler (1985) relation to the Aegean catalogue by Main and Burton (1989) that begins to attribute characteristic earthquake behaviour to shallow Aegean seismicity. TheLomnitz-Adler (1985) relation differs from the basic characteristic earthquake model of Youngs and Coppersmith (1985) in that instead of assuming uniform distribution for $M > M_{CH}$, a Gaussian distribution is assumed. A characteristic peak of Aegean earthquakes at M_S 7.0 ($\approx M_W$ 7.0) is identified for the back-arc extension of the Aegean and for the Hellenic arc, although only the back-arc events produce a statistically significant characteristic earthquake.

A more detailed analysis of seismicity along the Hellenic arc by Papadopoulos *et al.* (1993) would appear to confirm the characteristic earthquake model. Non-linearity of

the Gutenberg-Richter relation is apparent in most segments of the Hellenic arc. The exception to this is the Ionian island segment. In the investigation of the 2003 Lefkada earthquake Papadopoulos *et al.* (2003) present seismological and geological evidence for a characteristic earthquake along the Cephalonia transform fault too. The analysis by Papadopoulos *et al.* (1993) utilises the WWSSN catalogue in the period 1964 to 1985, and normalises this to the period 1901 to 1985. In retrospect, it is unclear to what extent the recurrence relation in each segment is biased by a single large event within a short time period. For example the characteristic for segment 1 (the Adriatic Coast) has a pronounced characteristic earthquake at M_S 7.1, yet the recurrence relation at higher magnitudes is dominated by one earthquake of that magnitude, occurring in 1979. Even allowing for the normalisation, the assertion of a characteristic earthquake in this segment may not be well-founded. As already noted, the Ionian segment does not appear to deviate from the G-R relation, though it may do so later. The SW Hellenic Arc displays no events above M_S 5.7 within the time period of the catalogue; arguments for a characteristic earthquake are based mostly on the September 13 1986 event (M_S 6.2). Again, a characteristic earthquake assertion is made based on one event. Similar results are apparent for the remaining segments.

It is not the intention of this discussion to refute the arguments presented by Papadopoulos *et al.* (1993), but one needs to be careful in applying more complicated recurrence relations in seismic hazard analysis. The examples of characteristic earthquakes being evident along the Hellenic arc, may indeed be an accurate reflection of the seismicity, but the model is scale dependent. Essentially the choice of zone and size of zone may control the significance of the characteristic earthquake.

Referring back to the cumulative magnitude-frequency plots of the declustered regional earthquake catalogue shown in Figure 2.9, it is evident that the characteristic earthquake identified is quite variable. For the pre-instrumental period there is little evidence to suggest a characteristic earthquake above the magnitude of completeness. However, for the early instrumental period a characteristic earthquake of approximately M_W 7 is visible. This is broadly consistent with previous estimates of the characteristic earthquake found in Main and Burton (1989). This is not entirely surprising as the events in the 1900 to 1963 catalogue comprise a significant component of the catalogues used in the earliest analyses of seismicity in the Aegean.

A characteristic earthquake is visible in the 1964 to 1998 period, yet from inspection it is less pronounced than for the early instrumental period. It is also closer to M_W 6.5, than to M_W 7.0. The final period (1999 - 2006) is poorly described by the Gutenberg-Richter relation, though this is largely due to the occurrence of the Izmit earthquake (M_W 7.6) within this brief time interval. Without the Izmit event it is very hard to identify any sort of characteristic earthquake.

When looking at the three catalogues terminating in 2005 (Subset 5, Subset 6 and Subset

7), it is clear that for the 1900 to 2000 and the 1964 to 2005 periods it is difficult to suggest that a characteristic earthquake model is preferable. The addition of historical events clearly elucidates the characteristic peak between M_W 6.8 and M_W 7. Yet the analysis of the historical period indicates that the most common earthquakes in that period have magnitudes in the range M_W 6.5 to M_W 7. The addition of historical events into the catalogue clearly has the effect of deviating from the fit from the Gutenberg-Richter relation, yet this is clearly a statistical artefact.

Ultimately, for the purposes of seismic hazard analysis a decision has to be made as to how to describe earthquake frequency magnitude recurrence. Although many authors have, in the past, made the case for a characteristic earthquake model and alternative recurrence relations, they shall not be implemented here. It is not entirely clear that even for smaller zones the recurrence is better described by relation with more parameters than those found in the truncated Gutenberg-Richter relation. Where such recurrence relations are fitted an additional source of uncertainty is added into the seismic hazard analysis. Efforts to incorporate both Gutenberg-Richter and characteristic earthquake models into a seismic hazard analysis by way of a logic tree have been applied in the USGS 2005 seismic hazard maps (Frankel *et al.*, 2005). The weightings for each branch appear subjective, and uncertainties in the parameters of each model are overlooked. It is felt that in the current hazard analysis for the Aegean, it is preferable to describe earthquake recurrence using the truncated Gutenberg-Richter relation, but allowing for the uncertainties in a-value, b-value and M_{MAX} to be explored within the Monte Carlo hazard analysis.

2.8 Minimum Magnitude (M_{MIN}) for Monte Carlo Seismic Hazard Analysis

2.8.1 The Necessity to Consider M_{MIN}

The minimum earthquake magnitude used in seismic hazard analysis is an often overlooked parameter. It is not uncommon to see minimum magnitude treated as an arbitrary value with little explanation given for ascribing that particular value. There are, of course, limitations to the earthquake catalogue that prevent small earthquakes being accurately recorded. This raises the question about the extent to which earthquakes below the magnitude of completeness should be included in seismic hazard analysis, since they are not represented in the earthquake catalogue. Consideration of these issues, and the sensitivity of seismic hazard to minimum magnitude (M_{MIN}), can be found in some references (Bender and Campbell, 1989; Grünthal and Wahlström, 2001; Beauval and Scotti, 2004), but this is small compared to the total number of seismic hazard studies completed to date.

The Monte Carlo method does have some advantages over the standard Cornell (1968) - McGuire (1976) methodology. Perhaps the greatest disadvantage, however, is the computational power required in its implementation. Whereas in the Cornell (1968) method, variation in M_{MIN} can be explored with relative ease, a decrease in the minimum magnitude in stochastic seismic hazard analysis can result in a dramatic increase in computer processing time. This is due to the increase in the number of earthquakes that are included in the synthetic catalogues, the resulting hazard from which has to be processed. The exponential nature of the Gutenberg and Richter (1944) recurrence relation means that for a region with b $\approx$ 1, a decrease in minimum magnitude of one magnitude unit will produce a ten-fold increase in the number of earthquakes required in the synthetic catalogue.

2.8.2 Considerations in Defining M_{MIN}

This brings us to the question of what must be taken into consideration when determining the minimum magnitude for seismic hazard analysis. Firstly the completeness of the earthquake catalogue (and the consequent fit of the recurrence relation) is an important consideration. Where the completeness magnitude is well below that magnitude at which damage is possible (another consideration that shall be addressed shortly) the hazard analysis can be sure that all the earthquakes are well represented in the catalogue. Otherwise, it is necessary to extrapolate the recurrence relation (typically the Gutenberg-Richter relation) to lower magnitudes and assume a predicted activity rate, as opposed to an observed activity rate. This of course means that activity rate (derived from the a-value) carries with it the uncertainty in the fit of the recurrence relation.

The second consideration is whether it is realistic to assume that all of the earthquakes with $M_W > M_{MIN}$ necessarily contribute to risk. The term risk being defined as the product of seismic hazard and the vulnerability of the structure(s) or location under consideration. Often the euphemistic term "engineering significance" is used to describe the ground motion that presents the possibility of damage to a structure, albeit this usually refers to engineered structures. Since damage correlates to several parameters of ground motion, in particular amplitude and duration of strong shaking; duration-dependent indices of strong motion often show a stronger correlation with damage than PGA or PGV. A particular threshold ground motion for engineering significance comes from Cumulative Absolute Velocity (CAV). A CAV of 160 cm s^{-1} is considered the threshold for damage in nuclear power plants (O'Hara and Jacobson, 1991; Cabañas *et al.*, 1997; Hardy *et al.*, 2006). This is determined using the CAV_5 (CAV integrated at a 5 cm s^{-1} threshold) attenuation relations of: 1) Kramer and Mitchell (2006):

$$\ln CAV_5 = 3.495 + 2.764\,(M_W - 6) + 8.539\left(M_W/6\right) + 1.008\ln\left(\sqrt{R_{JB}^2 + 6.155^2}\right) + \cdots$$
$$\cdots + 0.464F_N + 0.165F_R \pm 0.708P \qquad (2.27)$$

where R_{JB} is Joyner-Boore Distance, F_N and F_R are 1 for Normal and Reverse Faulting respectively and 0 otherwise; and 2) Danciu and Tselentis (2007):

$$\log_{10} CAV_5 = -1.665 + 1.138M_W - 2.304 \log_{10} \left(\sqrt{R_{EPI}^2 + 13.470^2} \right) + \cdots$$
$$\cdots + 0.063S + 0.234F \pm 0.595P \tag{2.28}$$

where S is 0, 1 or 2 for rock, stiff soil and soft soil respectively, and F is 1 for strike-slip and reverse faults and 0 otherwise. Using these attenuation relations, the range of magnitudes and distances across which expected CAV exceeds the 160 cm s^{-1} threshold can be estimated.

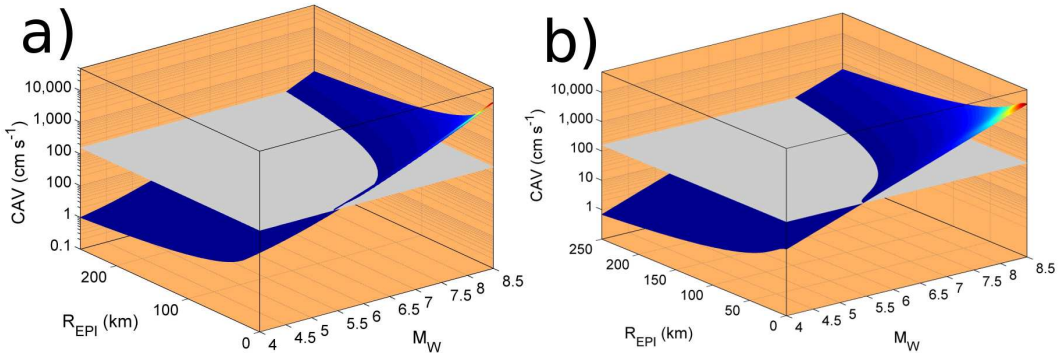

Figure 2.12: CAV attenuation (50th percentile) using the relation of a) Kramer and Mitchell (2006) and b) Danciu and Tselentis (2007). A normal faulting earthquake with 10 km focal depth is assumed. No soil dependence is assumed in the Kramer and Mitchell relation, soft soil is assumed for Danciu and Tselentis. The grey surface indicates the 160 cm s^{-1} threshold

The attenuation diagrams given in Figure 2.12 indicate that even in the extreme near field (epicentral distance < 1 km) the minimum magnitude needed to achieve this threshold is approximately M_W 5. This is lower than the completeness of the Aegean catalogue (1900 to 2005 A.D. - M_W 5.2), but higher than the completeness of the ISC/NOA period (1964 to 2005 A.D. - M_W 4.8). A M_{MIN} of M_W 5 would seem a reasonable choice for hazard analysis here. There are many years for which the catalogue is complete to this magnitude, against which it is possible to compare predicted activity rates to observed rates. However, the 160 $cm\ s^{-1}$ threshold is largely based on the performance of heavily engineered structures (Nuclear Power Plants). How does this compare with the damage that non-engineered domestic properties might experience? For this purpose macroseismic intensity is used as the shaking parameter, using the attenuation relation of Papazachos and Papazachou (1997) (Figure 2.13):

$$I = 1.43M - 3.59 \log (R_{EPI} + 6) + 2.26 \pm 0.87P \tag{2.29}$$

The preferred macroseismic intensity scale for Europe is the European Macroseismic In-

tensity Scale (EMS) (Grünthal, 1998). This is, for the purposes of attenuation, similar to the older Medvedev - Sponhaur - Karnik (MSK) scale, for which the Papazachos and Papazachou (1997) relation was derived. The intensity grade EMS/MSK 6 is shown in Figure 2.13, which is given as the damage threshold. At this intensity level many masonry buildings of vulnerability type A and B suffer grade 1 damage (superficial - fine cracks in plaster, some suffer grade 2 damage (non-structural cracks and partial collapse of chimneys). This is the threshold for anything other than superficial damage to the weakest (class A) buildings. It is at this level that there may be a risk to life and certainly an economic loss. As is seen in the attenuation relation, this level of damage typically occurs for earthquakes with $M_W \geq 5$. The threshold for any damage to class A buildings is EMS/MSK 5, which can occur for earthquakes as small as M_W 4. However, this is superficial damage and is highly unlikely to result in substantial economic loss. It should be noted that elsewhere in this study hazard at a site is only considered to occur from earthquakes within a 250 km radius. This is not an arbitrary choice, but arises from the limit of the damage radius ($EMS\,[MSK] \geq 6.0$) arising from the regional maximum earthquakes ($M_W \approx 8.6$).

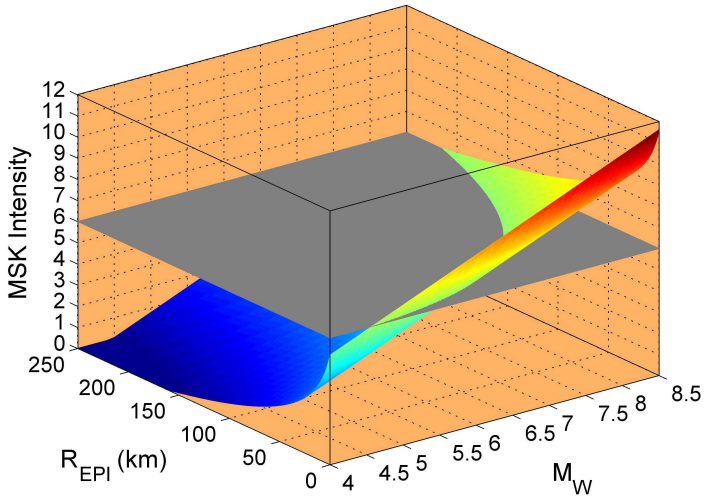

Figure 2.13: Intensity attenuation using the relation of Papazachos and Papazachou (1997)

The illustrations of attenuation used thus far have considered only the median (50th percentile) ground motion. Observed ground motions may deviate substantially from this relation, in a manner described by the log-normal scatter term found in the relation. It does mean that although the threshold levels of ground motion described are a useful guide, it is not to say that smaller earthquakes will not produce strong shaking above the threshold values. However, since the scatter is unbounded it is possible, though unlikely, for very small events to produce large ground motions if a sufficiently high number of standard deviations are used.

The third consideration in choosing minimum magnitude is the applicability range of the selected attenuation relation (see chapter 5 for further discussion). For example, the Danciu and Tselentis (2007) relation is derived from a data set spanning a magnitude

range of $4.5 \leq M_W \leq 6.9$. Extension of this relation to smaller magnitudes is a risk due to the non-linearity of ground motion. Equally, the strong motion threshold that is needed in selection of strong motion records for the derivation of the attenuation relation is such that there are few, if any, records from small earthquakes at larger distances. Other relations for the Aegean are simply not applicable for earthquakes with $M_W < 5$, with the exception of Bommer *et al.* (2007).

2.8.3 Impact of M_{MIN} on Seismic Hazard Analysis

Having discussed some of the important considerations in selecting M_{MIN}, it is worth noting the impact that variation in M_{MIN} has on seismic hazard. Sensitivity analyses performed by Grünthal and Wahlström (2001) and Beauval and Scotti (2004) revealed that lowering M_{MIN} increases the hazard for a site at short return periods, whilst having less impact on hazard for larger return periods. Disaggregation shown in Grünthal and Wahlström (2001) confirms this, with most hazard for short time intervals arising from small earthquakes in the near-field. This effect may be compounded if the preferred attenuation relation has magnitude dependent sigma (Ambraseys *et al.*, 2005a; Bommer *et al.*, 2007), in which case the hazard may come to be dominated almost entirely by small earthquakes.

It is not just the amplitude of strong motion that is sensitive to the choice of minimum magnitude; it is also the shape of the uniform hazard spectra. A more detailed description of this can be found in Bender and Campbell (1989), but the most important effect is that with lower minimum magnitudes, the higher frequency end of the spectrum increases to a greater degree than that at the longer period end. This is simply because smaller earthquakes tend to produce greater accelerations in the short period range than do larger magnitude earthquakes. When lowering M_{MIN} more small earthquakes are incorporated into the hazard and consequently will come to dominate the spectrum at higher probabilities of occurrence. The impact this will have from a seismic risk perspective, however, is not necessarily that great. Heavily engineered structures very rarely experience damage from high frequency strong motion, even without anti-seismic design. There may be a greater risk for small domestic properties, particularly those of vulnerability class A. Even then, the likelihood of structural damage or collapse remains small, and it may be only peripheral features of a building (such as the chimney) that would be damaged. Where there could be a potential risk is for machinery with particularly motion-sensitive components. Such machinery can be found in industry, power generation and hospitals, and could pose a potential economic risk.

2.8.4 Selection of M_{MIN} in this Analysis

The question remains as to what the choice of M_{MIN} should be for stochastic hazard analysis in the Aegean. The earthquake catalogue for the period 1900 to 2005 is complete

to approximately M_W 5.2. This spans a reasonable length of time and it is possible that, for the purposes of whole-process statistics, the true seismicity of the Aegean is well represented within this period of time, at least for small to moderate earthquakes. However, the 1964 - 2005 period is complete to M_W 4.8. This may be a sufficient length of time to allow for reasonable comparison of observed seismicity rate with that predicted by extrapolation of the Gutenberg-Richter relation to this magnitude.

Consideration of CAV and macroseismic attenuation relations would suggest that M_W 5 is not unreasonable as a lower bound magnitude. This is close to the magnitude of completeness, so it could be assumed that the additional earthquakes predicted from the Gutenberg-Richter relation for the range $5.0 \leq M_W \leq 5.2$ is a close approximation to the true result. This can be checked against the 40 years ISC section of the catalogue, which is complete to M_W 4.8.

For the Aegean region several acceleration attenuation relations have been identified that may be suitable (see Chapter 5). The Europe and Middle Eastern relation of Ambraseys *et al.* (1996) has a lower limit of M_S 4 ($\approx M_W$ 4.8), although regressions are performed for data with many lower magnitude limits. Skarlatoudis *et al.* (2003) and Danciu and Tselentis (2007) use M_W as the magnitude scale and are based exclusively on Greek earthquakes. The lower bound of the magnitude range of both these relations is M_W 4.5, whilst that of Ambraseys *et al.* (2005a) and Boore and Atkinson (2007) is M_W 5.0. Only the Bommer *et al.* (2007) relation may be applicable at lower magnitudes ($M_W \geq 3.0$). It can be implied from Figures 2.12 and 2.13 that earthquakes below magnitude M_W 4.0 pose little risk to property.

Several attenuation relations for duration dependent strong motion parameters are also considered, in addition to that of Danciu and Tselentis (2007). The CAV attenuation relation of Kramer and Mitchell (2006) has a lower limit of M_W 4.7, as does the Arias intensity attenuation relation of Travasarou *et al.* (2003). The lower bound of the Arias intensity relation presented in Kayen and Mitchell (1997) is unclear; however, there are few earthquakes smaller than M_W 6.1 in the data set used.

It is clear from these attenuations that for many relations, even those with M_S as the magnitude scale, the lower limit on magnitude is approximately M_W 5.0 (allowing for conversion between scales). More recent Greece-specific relations have a lower limit closer to M_W 4.5. None of the attenuation relations listed are believed to be suitable for earthquakes below this limit except Bommer *et al.* (2007). Furthermore it has been argued that not only should attenuation relations not be extrapolated beyond their applicability range, they may not even be accurate at the ends of their applicability range.

Although no magnitude range is listed for the macroseismic intensity attenuation relation of Papazachos and Papazachou (1997), the limitations on the magnitude range of other relations may suggest that it may not be safe to extend this relation to magnitudes

less than M_W 5. This is especially true when considering intensity in the near field of a small earthquake. The spectral content of waves emerging from a nearby small event is different to those felt at a distance from a larger earthquake. Specifically the nearby waves are richer in higher frequencies; consequently the macroseismic phenomena and damage observed may be fundamentally different.

What is apparent from this discussion is that there are several values of M_{MIN} that could be considered for the stochastic seismic hazard analysis: M_W 5.2 (the completeness magnitude for the 1900 - 2005 period), M_W 5 (onset of damage ($CAV \geq 160\ cm\ s^{-1}$, $EMS\,[MSK] \geq VI$, lower limit for older attenuation relations) and M_W 4.8 (lower 'safe' limit for Greek attenuation relations, and completeness magnitude for the 1964 - 2005 period). The intuitive way to explore the impact of M_{MIN} in the context of seismic hazard analysis might be to include M_{MIN} as a node of a logic tree. This approach is not adopted by Grünthal and Wahlström (2001), who instead prefer to consider different values of M_{MIN} in separate seismic hazard analyses. In this study, the Monte Carlo method is being used to analyse epistemic uncertainty. Since this will require simulating seismic hazard using catalogues that effectively sample alternative source models and attenuation models, it is preferable to consider variation in M_{MIN} separately, rather than within the epistemic uncertainty. This will make the impact of varying M_{MIN} immediately clear, rather than incorporating it into the many sources of epistemic uncertainty.

2.9 Conclusions

This chapter has introduced the earthquake catalogues that will be used throughout this thesis. In addition, analysis has been undertaken to define several key aspects of the seismic hazard model. This includes minimum and maximum magnitudes, magnitude-frequency relation, completeness magnitudes and declustering. Spatial and temporal variation of these parameters across the Aegean region has been analysed and interpreted.

It is clear that many of these parameters vary with scale, and so far only the region-wide catalogue has been considered. The next chapters will define seismic source models, which will require calculation of many of these parameters on a zone-by-zone basis. Decisions made in the delineation of source zones require that consideration be given to b-value, maximum magnitude and completeness magnitude. Furthermore, delineation of source zones via an algorithmic approach (Chapter 4) will also require that the data analysed in this chapter be used in a full seismic hazard analysis.

Chapter 3

Modelling Seismic Sources in the Aegean

3.1 Introduction

The development of a seismic source model is one of the most subjective and controversial elements of seismic hazard analysis. There exists little consensus on how best to approach the task, and what weighting particular types of data should be given. This problem is complicated further when addressing the issue of data quality, particularly with regard to earthquake catalogues. The uncertainty inherent in seismic source zonation, although smaller than the intrinsic scatter in strong ground motion relations, is far from trivial when applied to probabilistic seismic hazard analysis (PSHA).

Seismic sources may often be defined as points, lines or areal zones. When implemented in seismic hazard analysis, the epicentre of an earthquake may occur at any location along the line or within the zone, with uniform probability (Kramer, 1996). Hence, they are considered to be uniform zones. This assumption may be relaxed, as it is in the SEISRISK III software Bender and Perkins (1987), to allow for zone boundaries to be softened. In this approach the probability of an earthquake occurring at a location in the zone decreases away from the centre of the zone in a Gaussian manner. The exact definition of a seismic source (point, line or zone) in a particular region may depend not only on the information available, but on the nature of the seismicity itself. In low-seismicity intraplate regions all but the largest earthquakes are small enough to be approximated by a point source. In many cases it is not possible to attribute previous earthquakes in an area to a particular fault, in which case it becomes necessary to define a uniform zone. In more active areas coseismic slip may be observable on a particular fault; hence a line source approximation could be made. As a matter of conservatism, it may still be necessary to define an areal source zone around the line source to allow for uncertainty in epicentral location, and

also for the possibility of activity on smaller faults branching from the main structure.

In addition to the limits of available information and uncertainty in the location of future earthquakes, the context in which the seismic hazard analysis is being undertaken can influence the definition of the seismic source. For site-specific hazard analyses, the influence of fault dimension and behaviour will have a great impact on the ground motion estimated for the site. In which case, the seismic source is best defined by accurate determination of the physical dimensions of known faults in the area. For the purposes of mapping seismic hazard this approach is not usually possible, and the characteristics of individual structures will have little impact on the seismic hazard once spatial interpolation is applied. In this case, spatial zones may prove to be the most suitable method of characterising the source.

For many countries where seismic activity poses a threat to persons and property, seismic hazard maps are regularly updated. It is common practice for the seismic source models (the collection of source zones across the region) to be updated with them. A good example of a source model can be seen in the recent seismic hazard maps for Italy (Montaldo *et al.* (2005) - Figure 3.1).

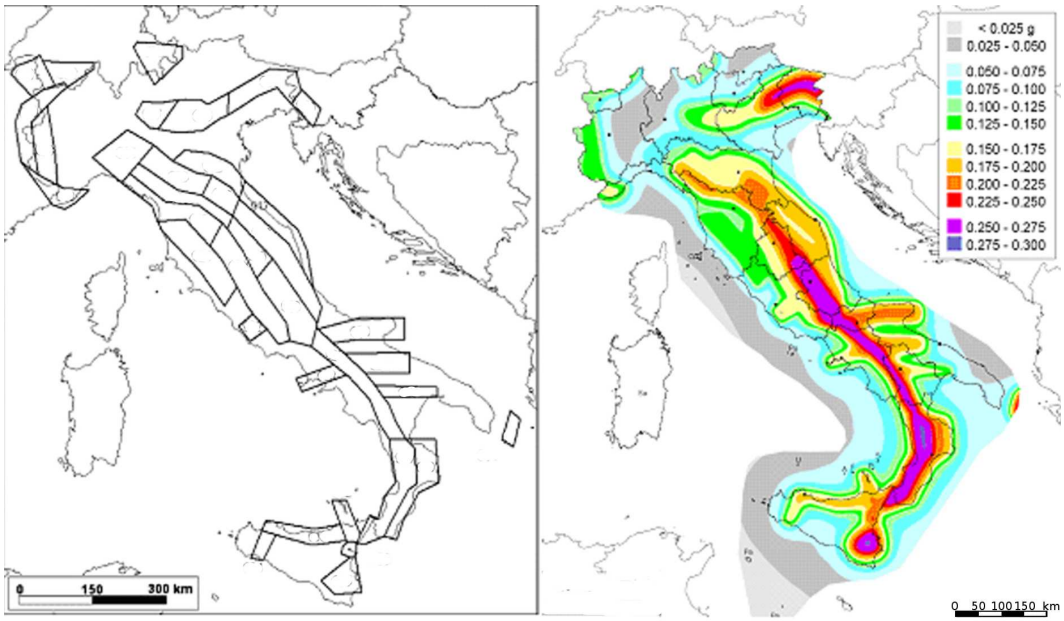

Figure 3.1: a) Seismic source model for current seismic hazard maps of Italy (Montaldo *et al.*, 2005), b) Seismic Hazard map in terms of PGA with a 10 % probability of being exceeded in 50 years on a rock site (INGV, 2005)

When producing a seismic source model such as that shown in Figure 3.1, there is often no formal guidance or methodology as to how to delineate seismic sources. An advantage of this is that is it easier to update source models over time as new information becomes available. It is common for source models to be constructed by a panel of experts, who take into consideration geology, seismotectonics and seismicity, weighted according to their own judgement. The results are often subjective; the result of a compromise of ex-

perts rather than necessarily each expert's best judgement. Despite this, the use of expert panels is still perhaps the best way to incorporate such a wide variety of information into the model. An attempt to automate the process of assimilating information from various geophysical and geological sources, thereby providing an objective source zonation, has been attempted by Zamani and Hashemi (2004) using cluster analysis. Whilst the results were of interest, the manner in which geophysical data are treated in the cluster analysis may still be considered subjective. A novel method using K-means cluster analysis will be presented in the next chapter.

Alongside the issue of subjectivity in the development of the seismic source model, is the lack of objectivity in assessing the quality of the model. This is frequently done by expert judgement. Quantitative approaches to assessment of a source model's quality have been suggested (Musson, 2004), where judgement is usually based on the ability of a model to replicate observed seismicity. This approach will be extended further in the next chapter by comparing the ability of a source model to replicate observed hazard. Even these quantitative approaches may contain bias, depending on the relative role that seismicity has played in the models delineation. Observed seismicity primarily guides the delineation of the seismic source, be it point, line or zone, then this source model will naturally provide a better fit to observed seismicity than a model derived using alternative geological and seismotectonic information. More discussion of the assessment of source model quality will be given in later chapters.

3.2 Zone-Free Methodologies

The innate error and subjectivity in the zoned approach to seismic source modelling has led to the search for a zone-free methodology. This can take many forms. Several popular techniques have emerged, which tend to rely on observed seismicity almost exclusively for information. As a result, they are often applied in conjunction with a fixed source zone approach

One of the most widely used zone-free approaches is the spatial smoothing method (Frankel, 1995; Frankel *et al.*, 1996). This requires the division of the study region into a uniformly spaced grid, counting the number of earthquakes with $M > M_C$ (where M_C is the completeness magnitude) in each grid cell. The logarithm of this number represents the maximum likelihood a-value for the cell. The a-values are then smoothed using a Gaussian function with a correlation distance specified by the user. Frankel *et al.* (1996) demonstrate that the choice of correlation distance has a significant impact on hazard, yet in practise the choice of this parameter is often quite arbitrary. Furthermore, the entire area being analysed requires the definition of a single b-value and maximum magnitude. The necessity of assigning single recurrence parameters to a large region means that this method is not used where seismicity can be clearly attributed to known sources. It is still applicable as a way of describing background seismicity, or seismicity in intra-plate

regions where seismogenic sources are ill defined.

The Frankel (1995) approach can be adopted for hazard assessment using the Cornell (1968) and McGuire (1976) methodology. Other approaches undertake the hazard analysis via alternative means, again using the observed spatial distribution of earthquakes as the input for the hazard assessment. The simplest of these approaches is the Historic-Parametric method Bommer *et al.* (1998). When source dimensions are small compared to source-site distances, ground motion at a site follows the following power law relationship:

$$\log(N) = \alpha_1 + \alpha_2 \log(Y) \tag{3.1}$$

N is the number of events per year capable of causing ground-motion value Y, from all seismogenic sources, to be exceeded. α_1 and α_2 are constants determined by regression. To determine N, one simply calculates the expected ground motion at a site from every earthquake in the catalogue and fits the distribution in 3.1 to the ground motions. Design values can be simply extracted from this distribution.

An amalgam of the spatial smoothing method and the historic-parametric method is the kernel estimation technique of Woo (1996). This approach arises from the observation that the fractal dimension (D) of earthquake epicentres has a dependence on magnitude. Here the rate of seismic activity (λ) at a site x is estimated via a statistical smoothing operation, using a Kernel function (K):

$$\lambda(M, x) = \sum_{i=1}^{N} \frac{K(M, x - x_i)}{T(x_i)} \tag{3.2}$$

M is the magnitude of the earthquake with epicentre x_i and effective observation period T(x_i), which is an element of the catalogue of N events. Various Kernel functions can be found in existing literature (see references in Woo (1996)) and have been applied to seismicity. Using this method it is possible to form a discrete grid of activity rates in a similar fashion to the Frankel (1995) approach.

An alternative to the Parametric-Historic methodology is the Extreme-Value approach (Makropoulos and Burton, 1985b; Al Abbasi and Fahmi, 1991; Burton *et al.*, 2003, 2004b). This can take one of two forms: calculation of design strong ground motion exclusively, or calculation of design magnitudes (as seen in Chapter 2). Both approaches begin by considering only the seismicity within a moving grid cell of a fixed size (square or circular - square being preferred), centred on a site. If using strong motion as the parameter, an appropriate attenuation relation is used to calculate strong motion at the site from each earthquake selected. The largest ground motion within a set time period is then taken as the extreme value for input into the extreme value data set. This data set is then sorted into numerical order and the probability of each annual maximum is calculated using the Gringorten (1963) plotting probability formula. A Gumbel I distribution is then fitted

(via linear least-squares) to the probabilities:

$$G^I(y) = \exp(-\exp(-\alpha(y-u))) \tag{3.3}$$

Where α and u are parameters of the distribution and $G^I(y)$ is the probability of y being an annual maximum ground motion. The Gumbel extreme value approach can be sensitive to a number of input parameters. The size of the cell will influence the hazard calculation, and hazard maps will be influenced by the resolution of the moving cell, which will create an artificial smoothing effect.

The Monte Carlo approach of Ebel and Kafka (1999) can also be used as a zone-free method of estimating seismic hazard. Like the previous methods it utilises direct calculations of ground motion at a site using a selected attenuation relation. The main difference is that whereas the Historic-Parametric method fits a distribution to the ground motion arising from observed earthquakes, the Monte Carlo approach uses repeated simulations of seismicity (synthetic catalogues). The design hazard is determined by selecting the ground motion level (for a site) that is exceeded a required number of times in the synthetic catalogue. The similarity between this and the Historic-Parametric method is that both use the observed seismicity to define discrete point sources. The synthetic catalogues produced by Ebel and Kafka (1999), capture spatial distribution of seismicity by randomly re-sampling, without replacement, the observed epicentres. Rather than using the sampled epicentres directly, the synthetic epicentres are actually sampled from a Gaussian distribution of epicentral location, which is centred on the observed epicentre. This method is again, most suited for regions where the physical sources are ill-defined, such as in the Eastern United States where the methodology is applied in the aforementioned reference.

The common thread that exists in these methodologies is that they are most appropriate for regions where the seismic sources are constrained almost exclusively by observed seismicity. In many of the references quoted, the illustrative application has been to intraplate regions e.g. Central United States (Frankel *et al.*, 1996), Eastern United States (Ebel and Kafka, 1999), United Kingdom (Woo, 1996). Where seismogenic faults are clearly defined and their behaviour adequately modelled, the spatial smoothing methods suggested may not be appropriate. In fact, application of these methods could produce an excessive smoothing of activity rate, resulting in underestimation of hazard along highly active structures. Hazard maps produced by these methods frequently display an abundance of isolated maxima and minima of strong ground motion along plate margins. This is because of the powerful influence of spatial clusters of earthquakes in the seismic catalogue being used. Furthermore, the application of spatial smoothing methods to low or moderate seismicity regions is made tolerable by the prevalence of lower magnitude earthquakes ($M < 6.0$). For smaller earthquakes, the approximation to a single point source is not unreasonable given the distance scales involved in many hazard analyses. However, large earthquakes have an exponentially longer rupture length; the assumption

of a point source is no longer valid. Frankel *et al.* (1996) tend to overcome this problem by introducing small uniform source zones in regions where larger earthquakes have occurred in the past.

3.3 Linear Sources and Hybrid Models

Where seismogenic faults have been well-constrained by geophysical and geodetic investigation, and seismicity, the option of defining a linear source is available. Given our understanding of the nature of fault rupture, this would appear to be the optimal solution. In reality, the seismogenic properties of a fault are rarely consistent with the model of a single planar surface rupturing independently of surrounding tectonic structures. The interdependence of earthquakes via stress transfer (King *et al.*, 1994), and the fractal nature of fault systems (Turcotte, 1997), means that branch faults and surrounding fractures will be strongly influenced by movement on the main fault. Furthermore, there are very few locations where faults systems are sufficiently well-mapped to allow for identification of all active seismic sources. Even within these regions, such as California or Japan, it is still very difficult to characterise, adequately, behaviour to the extent necessary for seismic hazard analysis.

The compromise to this problem is what is commonly practiced in regional seismic hazard analyses across the globe. This is often termed the "hybrid approach", and it requires the definition of active fault sources where possible, and seismic source zones beyond the limits of the fault. This may be considered an optimal-information method, whereby the influence of well defined sources will take precedence in the active zones. At the same time, where seismic sources are poorly defined, the activity is characterised by a zone of uniform Poissonian seismicity. Whether or not this should be defined as "background" seismicity can be a subjective choice, which is specific to the region and scale of the seismic hazard analysis being conducted.

3.4 Previous Seismic Source Models of the Aegean Region

The definition of seismic sources in the Aegean region is a complicated one, yet there have been many consistencies in the zone models developed in the last 30 years. The high seismic activity of the Aegean has made seismic source modelling a topic of paramount importance in European seismic hazard analysis. In addition to the models developed specifically for Greece, some parts of the study area are covered by zones pertaining to the source models of neighbouring regions (e.g. Adriatic zone (Slejko *et al.*, 1999), Turkey (Erdik *et al.*, 1999), the Balkans (Musson, 1999b; Ardeleanu *et al.*, 2005)).

3.4.1 Hatzidimitriou *et al.* (1985) - HZ1985

The earliest source model with defined seismicity parameters considered here is that of Hatzidimitriou *et al.* (1985), hereafter referred to as HZ1985. This model defines 21 zones across Greece, Albania, Macedonia, Southern Bulgaria and Western Turkey (Figure 3.2). Generally these zones fall into three regions, distinguished by variation in b-value. Region A corresponds to the outer arc, with b-values in the range 0.88 to 1.23 (with an average of b = 1.03 $\pm$ 0.11). Region B corresponds to the back arc and has lower b-values (0.79 to 0.92, with a mean of 0.84). Region C corresponds to the Balkan and Western Turkey region, believed to be a more structurally stable area, and has lower b-values still (0.49 to 0.71, with a mean of 0.60). A case is made for the correlation of b-value to geology from southwest to northeast across the Aegean. This is done on the basis of geological heterogeneity, with older crystalline Palaeozoic and Pre-Paleozoic rock of the east Balkan region considered more structurally stable than the Mesozoic and Tertiary sediments that characterise much of the external Hellenide belt.

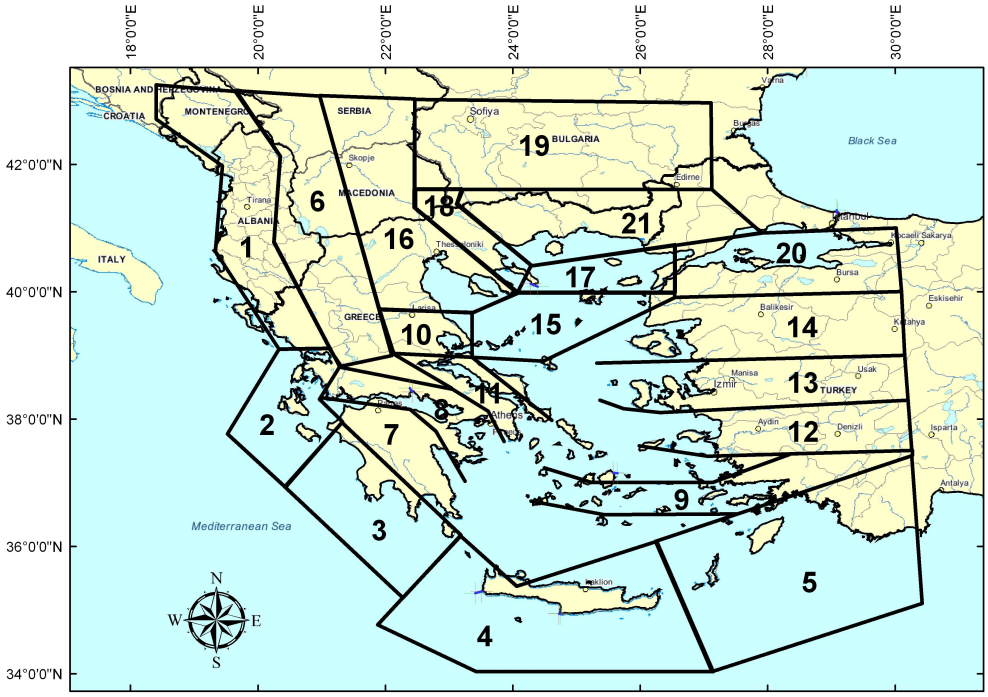

Figure 3.2: HZ1985 shallow source model (Hatzidimitriou *et al.*, 1985)

The comparative role of geology and tectonic activity in the Aegean is not well addressed by this model. The authors do recognise that b-value can vary with time in a region. Whilst certainly an important step in the beginnings of detail seismic hazard analysis in the Aegean, the zones presented in HZ1985 have limitations in their use as a source model for this purpose. Several zones in the model are open-ended in the low seismicity region of the South Aegean Sea. This makes hazard in the south Aegean region harder to define, a particularly great problem given the historical earthquake damage to Athens.

3.4.2 Papazachos (1990) - PZ1990

The next model, and arguably the first to be applied in seismic hazard analysis in Greece, is that of Papazachos (1990), hereafter referred to as PZ1990 (Figure 3.3). This model contains 36 uniform zones encompassing the same area as that of HZ1985. Areas outside these zones were considered to be background seismicity - ostensibly uniform. This includes the south Aegean, Gulf of Patras and much of Macedonia.

Papazachos (1990) does present a more detailed analysis of the definition of each zone within this model than Hatzidimitriou *et al.* (1985). Clear justifications for the boundaries of each zone are presented, which makes scrutiny much easier. The role that each source of information (seismicity, geology, focal mechanism) plays is apparent, and in that respect there exists a solid basis for establishing particular zones. Nevertheless, there is still reason to suggest that some zones were subdivided either with reluctance or on the basis of the supply of information rather than on the basis of what the information shows. This is particularly true of offshore zones where seismic completeness can vary considerably.

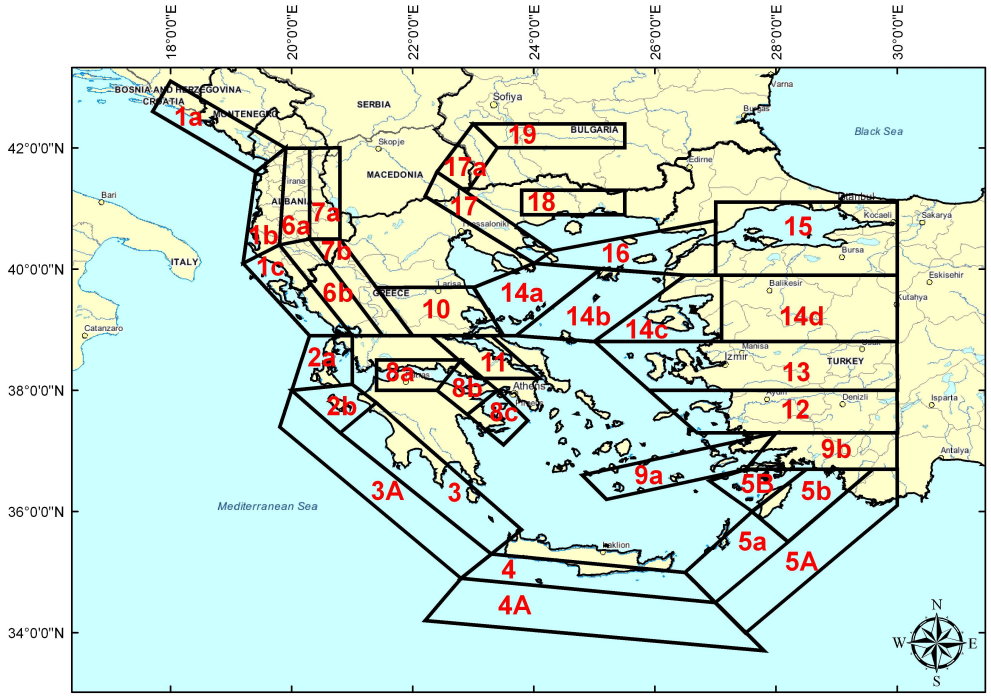

Figure 3.3: PZ1990 Shallow Seismic Source model (Papazachos, 1990)

The definition of zone models in the context of seismic hazard analysis is addressed within this model. Earthquake recurrence parameters are listed for the 36 zones, yet b-values are simply divided into three classes: b = 0.6, 0.8 or 1.0. The decision to assign arbitrary b-values could very well be to avoid assigning extraordinarily high or low values because of the transient seismicity within small zones: a problem that is addressed differently in later models. Given the variation in tectonics across even stable regions, this may also be a spurious approach which masks some of the real variation in earthquake

behaviour.

Perhaps the greatest concern with this model is the designation of "background" seismicity. Of particular concern is that a large area of Northern Greece, an area that has experienced some large historical earthquakes, falls within this region. Background seismicity is not an original concept, but it is may be a controversial one, especially in an active part of the world such as the Aegean. Such a designation is more tolerable for offshore regions, yet onshore there exists a risk to persons and property. The background region in the PZ1990 model is assigned a b-value of 0.8 and M_{MAX} of M_W 6.1. Justification for this comes from the observation that any earthquakes of M_W 6.1 or greater are expected to have very long return periods. The decision to designate background seismicity was of course criticised following the Kozani-Grevena earthquake of 13 May 1995 (M_W 6.5). This was a particularly damaging shallow earthquake occurring in Northern Greece, within the "background" region.

Papazachos (1990) does make an initial attempt to designate zones of intermediate depth seismicity along the Hellenic arc. Here, two arc-shaped zones are each split into three segments from east to west. It is known that some of the largest recorded earthquakes have occurred at depth along the Benioff zone of the Hellenic arc, which presents a hazard for engineered structures in the Aegean. The two arcs occur at different depths, with the outer arc (Intermediate zones 1a, 1b and 1c) covering the range 70 to 100 km, and the inner arc (Intermediate zones 2a, 2b and 2c) 100 to 160 km. The outer arc displays more seismic activity due to coupling of the subducting east Mediterranean lithosphere with the more buoyant Aegean lithosphere. It is here that the larger earthquakes have occurred. The angle of dip for the subducting lithosphere in the outer arc has been shown to be close to 10°. The inner arc displays lower seismicity and dips more steeply ($\approx 38^{\circ}$). The lower seismicity here suggests decoupling of the two lithospheric slabs. The prevalence of large earthquakes results in the outer arc (zones 1a, 1b and 1c) having a substantially lower b-value than elsewhere in the Aegean, calculated as 0.56. The inner arc (zones 2a, 2b and 2c) has a slightly higher b-value of 0.75. There is relatively little variation in a-value and M_{MAX} within the inner and outer arc, thus leading one to question whether the partitioning into three zones is entirely necessary.

3.4.3 Papaioannou and Papazachos (2000) - PP2000

Further research, and the occurrence of the 1995 Kozani-Grevena earthquake, prompted redevelopment of the seismic source model of the Aegean, which resulted in the model formalised by Papaioannou and Papazachos (2000), initially presented in Papazachos and Papazachou (1997) and hereafter referred to as PP2000. This is a 67-zone continuous model that spans the entire Aegean region (Figure 3.4). An additional seven intermediate depth zones are also defined. Basic seismicity parameters for these zones, as well as expected isoseismal azimuth and ellipticity, are given in the original source.

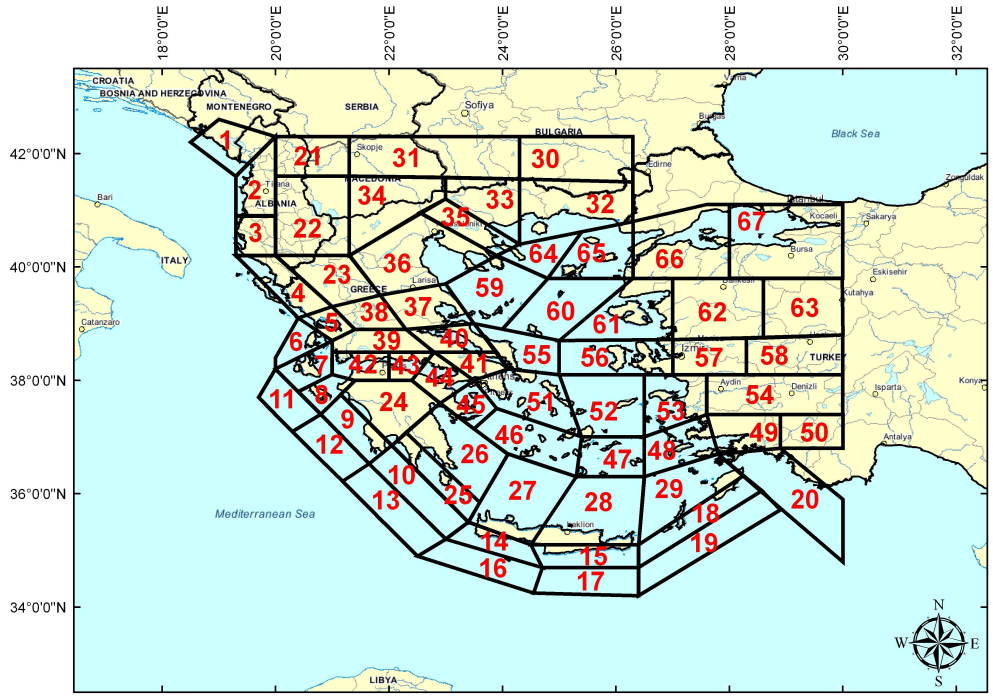

Figure 3.4: PP2000 Shallow Source model (Papaioannou and Papazachos, 2000)

The PP2000 model has, until recently, been considered the standard seismic source model for Greece. The obvious feature that distinguished this model from those previously (PZ1990 and HZ1985) is the absence of a background zone. Now, almost all of Greece, Albania, Macedonia and Western Turkey are covered by the zone model. The replacement of a "background" with uniform zones has clearly resulted in an increase in the number of source zones used. There are many regions where PP2000 source zones and PZ1990 zones appear very similar. This result is unsurprising. In many locations the PZ1990 zones have been subdivided further, or in some cases amalgamated.

Unlike Papazachos (1990), neither Papazachos and Papazachou (1997) nor Papaioannou and Papazachos (2000) provide an explanation as to how the PP2000 zones were determined. Where there is strong similarity between the PP2000 zones and PZ1990 zones, it is not unreasonable to expect that justification to have been carried forward into the new model. It is zonation in the areas previously ascribed to be background that provide some perplexing decisions. In these regions many of the previous underlying assumptions still hold true. It is not clear on what basis the source zones were determined in PP2000, when they were left unzoned in PZ1990.

The seismicity parameters determined for the PP2000 model, also represent a departure from previous estimates. Instead of estimation by simple least squares, a- and b-values have been determined using the linear smoothing method of (Papazachos, 1999a). This method requires that for $i = 1 \ldots K$ cells, the complete catalogue of earthquakes for each cell can be assigned to j magnitude - cumulative number pairs ($[N_{ij}, M_{ij}], j = 1, \ldots, n_i$),

resulting in the equation:

$$\log N_{ij} = a_i + b_i M_{ij} = \sum_{k=1}^{K} \delta_{ik} \left(a_k + b_k M_{kj}\right) \tag{3.4}$$

where δ_{ik} is the Kronecker delta. This is then formulated as a linear system:

$$\mathbf{N} \;=\; \mathbf{A} \left[\begin{array}{c} a \\ b \end{array} \right] \;=\; \mathbf{Ax} \tag{3.5}$$

where N contains the $\log\ N_{ij}$ values, a and b are the vectors containing the a_i and b_i values for the respective cells. Smoothing arises from the constraints $\lambda\partial^2\mathbf{b} = 0$, $\mu\partial^2\mathbf{a} = 0$ and $\upsilon\partial^2\tilde{\mathbf{a}} = 0$, where λ, μ and υ are smoothing factors (1, 2 and 1 respectively in Papazachos (1999a)), and $\tilde{\mathbf{a}}$ the vector of cells containing too few earthquakes to accurately estimate seismicity parameters ($n_i \;<\; n_{cut}$). If a and b are scaled to their *a priori* uncertainties this produces a covariance matrix, resulting in the system:

$$\mathbf{C_N^{-1/2}N} = \mathbf{C_N^{-1/2}Ax}, \tag{3.6}$$

$$\lambda\partial^2 \left[\mathbf{C_b^{-1/2}b}\right] = 0, \tag{3.7}$$

$$\mu\partial^2 \left[C_a^{-1/2}a\right] = 0, \tag{3.8}$$

$$\nu \left[\mathbf{C}_a^{-1/2}\tilde{a}\right] = 0 \tag{3.9}$$

This system is solved using least squares to produce the vectors **a** and **b**.

This smoothing approach has the impact of narrowing the range of b-values to 0.79 to 1.02. Most zones fall into an even narrower range, with many neighbouring zones having b-values differing by less than 0.05. Papazachos (1999a) implements this method due to the recognition that large variation in a- and b-value can emerge as a result of numerical instabilities in the data, rather than true spatial variation. Therefore to understand how b-values vary spatially across a region due to physical variation in seismotectonics a smoothing approach is necessary.

Underlying the philosophy of smoothing is the assumption that spatial variation in b-value is determined for the purposes of geophysical investigation, not necessarily for the purposes of hazard analysis. In the latter, it is imperative that the uncertainty on b-value be known and incorporated into the hazard calculations. In this case, smoothing the seismicity parameters may mask pertinent variation in seismicity, which could have implications for hazard analysis. Furthermore, where neighbouring zones display near identical b-values, and other parameters such as a-value and M_{MAX} are derived from these, then it may be the case that for hazard analysis the zones are amalgamated, albeit allowing for an increase in the uncertainty of the parameters. Whilst the 67-zone model of PP2000 improves the spatial resolution of seismicity variation, the smoothing approach may then diminish this variation. This leads to the question of whether 67 zones is too

many for the purposes of practical seismic hazard analysis.

Intermediate depth source zones are also presented in the PP2000 model. These generally formalise the speculative zones presented in PZ1990. The only significant modification is that the outer arc zone, which previously contained three sub-sections, has now been partitioned into four sub-sections. Estimates of b-value for the inner and outer intermediate zones are the same as those found in PZ1990 (i.e. 0.75 and 0.56 respectively). Maximum magnitude estimates for the intermediate depth zones have been generally revised downward.

3.4.4 Erdik *et al.* (1999) - EK1999

Though not explicitly designed for the purposes of PSHA for the Aegean region, the 37-zone seismic source model for Turkey also covers much of the Aegean. Zones 1 to 25 span Greece, the Balkans and the Aegean Sea (Figure 3.5). The methodology behind the delineation of these zones is not made explicit; a general scheme is implemented. These included the following conditions:

- "Sources [zones] should be defined as areas with seismic characteristics that are as homogenous as possible."
- "Between sources of different seismic potential, the boundary should be located close to the highest concentration around the hard core of the most active ones."
- "In areas possessing a statistically significant number of reliable events, boundaries should be mainly based on seismic data as an expression of tectonic activity, and backed up by tectonic arguments."
- "In case of an insufficient number of events or a large number of uncertainties attached to these events, existence of a boundary has been decided by arguments based on the most dominant tectonic or seismic features."

Whilst these arguments do provide the basis for an approach to zonation (they are of course instinctive), the output is far from objective. Furthermore, this approach will result in the definition of "background" seismicity, the implications of which have been discussed in the PZ1990 model.

The resulting model of Erdik *et al.* (1999) does display some interesting features. In particular, zones in this model display a greater degree of curvature, following the shape of major fault systems rather than partitioning them. Large areas of similar tectonic homogeneity such as the Hellenic Arc, the Gulf of Corinth and the North Anatolian Fault are preserved as single zones. An implicit assumption within this approach, therefore, is that the variability in the seismic activity within these regions, in the observed catalogue, is a transient feature.

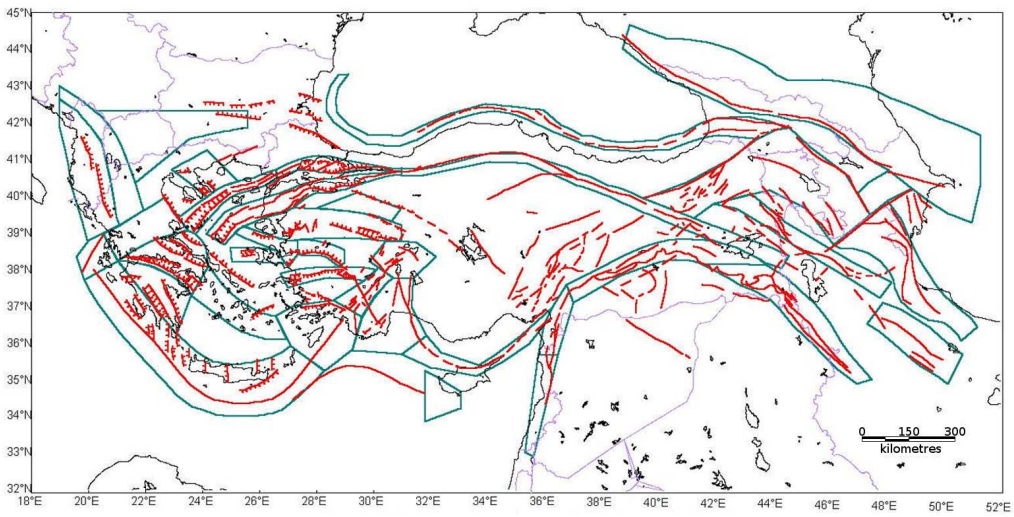

Figure 3.5: EK1999 Shallow seismic source model for Aegea and Anatolia (blue polygons) (Erdik *et al.*, 1999) with major tectonic features (red)

3.4.5 Jiménez *et al.* (2001) - JM2001

As part of the Global Seismic Hazard Assessment Program, it was necessary to produce a unified seismic source model to homogenise the sources across the European region (Jiménez *et al.*, 2001). For shallow seismicity in the Aegean region the PP2000 model was combined with the Erdik *et al.* (1999) model of Turkey (EK1999). Intermediate-depth source zones are taken from the PP2000 model.

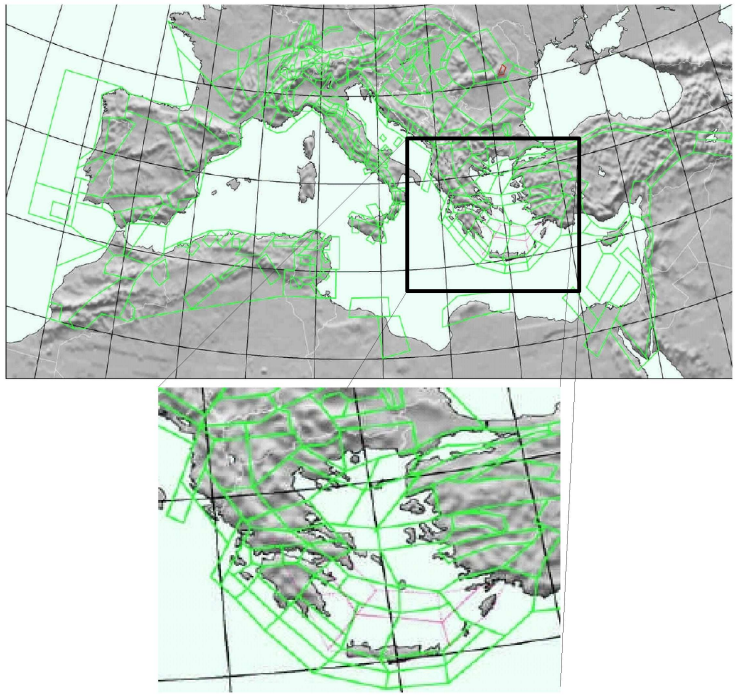

Figure 3.6: JM2001 Shallow source model for Europe, with the Aegean shown in detail (Jiménez *et al.*, 2001)

The method used to produce this European-wide homogenised model (Figure 3.6) is broken down into two steps. The first step is to identify existing models in such a manner as to ensure as great a spatial coverage as possible. These zones are then joined. Where zones from different models overlapped these were "redesigned to harmonise geometries where differences existed". Whilst the first step is clearly logical, in the absence of further detail the latter step is quite puzzling.

Despite the lack of transparency in the delineation of the source model, there are many features of JM2001 that could be considered a reasonable representation of the seismotectonics of the Aegean region. In particular, the combination of the PP2000 and EK1999 models in the Sea of Marmara and North Aegean Sea, results in a series of zones that are more tightly constrained to seismicity around the North Anatolian Fault and its extension into the North Aegean Sea. Similarly the reduction in the number of shallow zones, from 67 in PP2000 to 43 in JM2001, may address some of the concerns of PP2000 raised previously regarding the neighbouring zones with similar properties.

3.4.6 Papaioannou (2006) - PA2006

The most recent source model is that of Papaioannou (2006 - personal communication) (PA2006). This is a hybrid model where small-intermediate earthquakes are considered as point sources, occurring uniformly within each zone (Figure 3.7). Larger earthquakes are modelled as ruptures, occurring on discrete fault segments. This is a departure from the previous models as it is now beginning to distinguish between the assumptions of seismic source for both small and large events. This has the advantage of constraining the location of the largest earthquakes to known ruptures, whilst allowing for the uncertainty in location of smaller events.

Once again the number of zones is smaller than that of PP2000 with large areas of the Balkans and Western Turkey encompassed by single zones. Also seismicity in the Sea of Marmara and North Aegean Sea is more tightly constrained to the trace of the North Anatolian Fault. Many of the zones represent a radical change from the PP2000 model. For example the Gulf of Corinth is here treated as a single zone, as is the Thessalia region to the north. Previous models had divided these areas into smaller zones.

It remains to be seen how this model influences seismic hazard analyses in the Aegean region. A preliminary hazard map produced by Papaioannou (2006) is shown here (Figure 3.8). Highest hazard can now be seen in the Ionian Islands, the Gulf of Corinth, the North Aegean Sea and Sea of Marmara, as well as Southern Bulgaria.

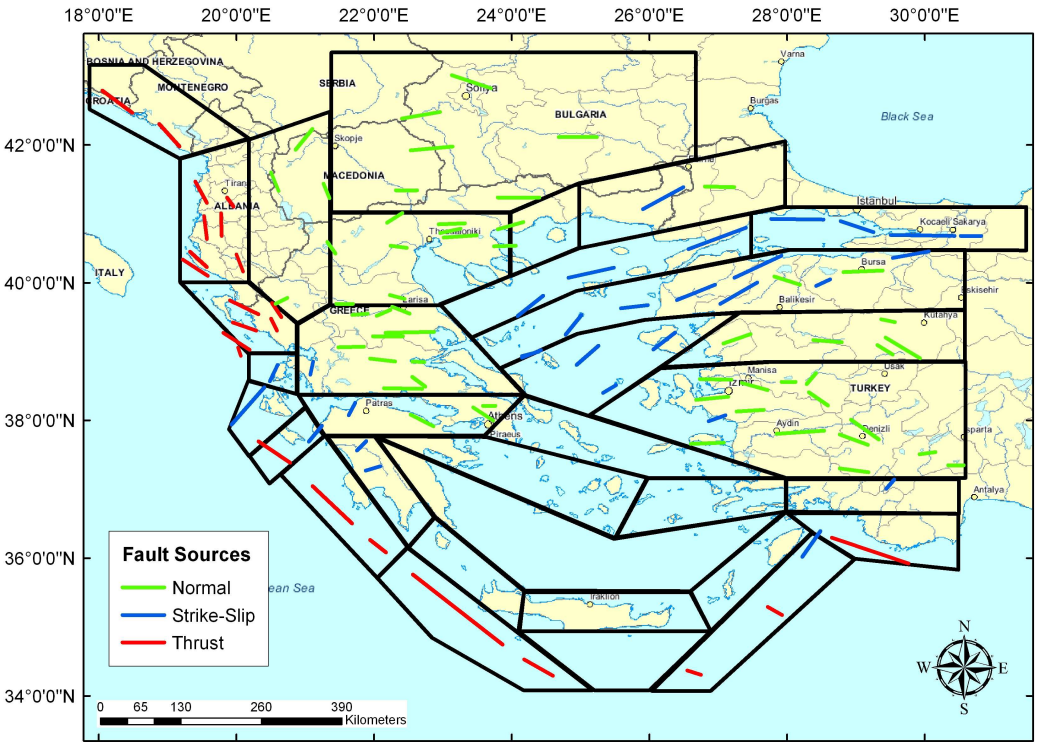

Figure 3.7: PA2006 Hybrid Shallow Source Model (Papaioannou, 2006 Personal Communication) with fault sources shown

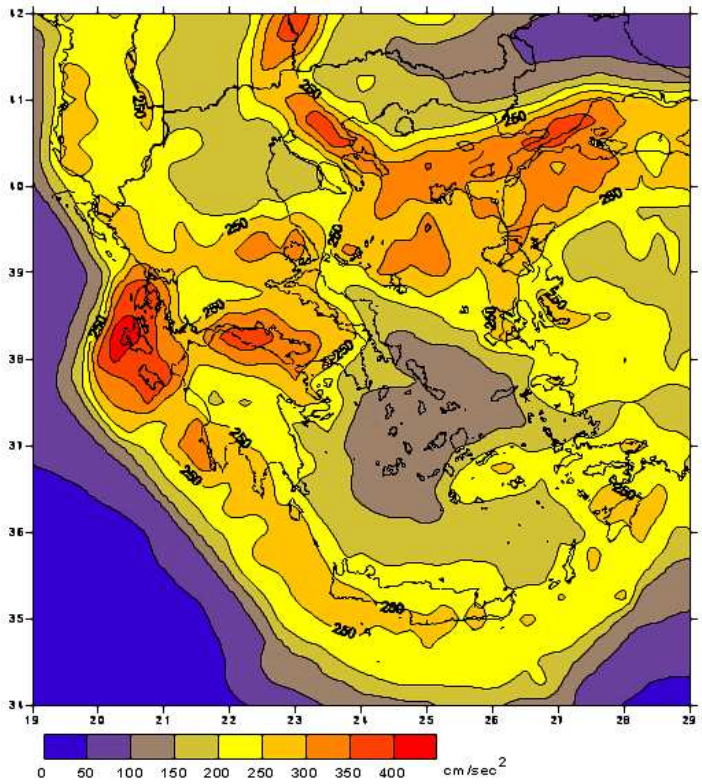

Figure 3.8: Map of PGA with a 10 % probability of being exceeded in 50 years, using the PA2006 hybrid source model (Papaioannou, 2006 Personal Communication)

3.4.7 Other Models

Being the most seismically active region in Europe, parts of the Aegean are often covered in source models for other areas. The seismogenic source model of Slejko *et al.* (1999) for the Adriatic region is one such example (Figure 3.9). This model contains source zones pertaining to seismicity in Montenegro, Albania, Northern Greece and the Ionian Islands, regions that are also covered by Aegean source models. Generally the ones in the Slejko *et al.* (1999) scheme cover only the Adriatic coast, and hence do not penetrate too far inland. In this region the zones generally take the form of simple parallelograms that follow the trace of the Africa-Eurasia boundary as it extends along the Adriatic coast. Parallel to these zones is another line of source zones, which are designated as background seismicity. This is to allow for the treatment of seismicity outside the study area in the same computational procedure as the "active" zones. Some justifications are given for this zoning model, especially where the zonation may appear controversial, such as in the Ionian Islands and western Albania.

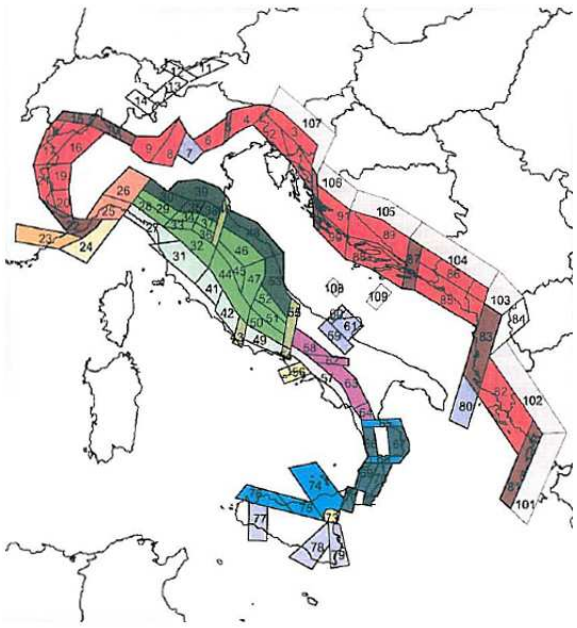

Figure 3.9: Model of Seismic sources for the circum-Adriatic region (Slejko *et al.*, 1999). Colours indicate tectonic environment (refer to Slejko *et al.* (1999) for details)

Much like the Slejko *et al.* (1999) model, the seismic hazard analysis for Bulgaria and Romania of Ardeleanu *et al.* (2005) includes seismogenic zones in southern Bulgaria that can be compared with those of Aegean models (Figure 3.10). Here just two zones span southern Bulgaria, with a third covering the Macedonia-Serbia border region. Little information is given to justify this particular zonation. The region in question is characterised by a low strain rate and low seismicity, typical of an intraplate region.

In addition to the seismotectonic models, Koravos *et al.* (2003b) and Koravos *et al.* (2003a) utilise a 16-zone model based on geodetic information (Figure 3.11). Within each of these

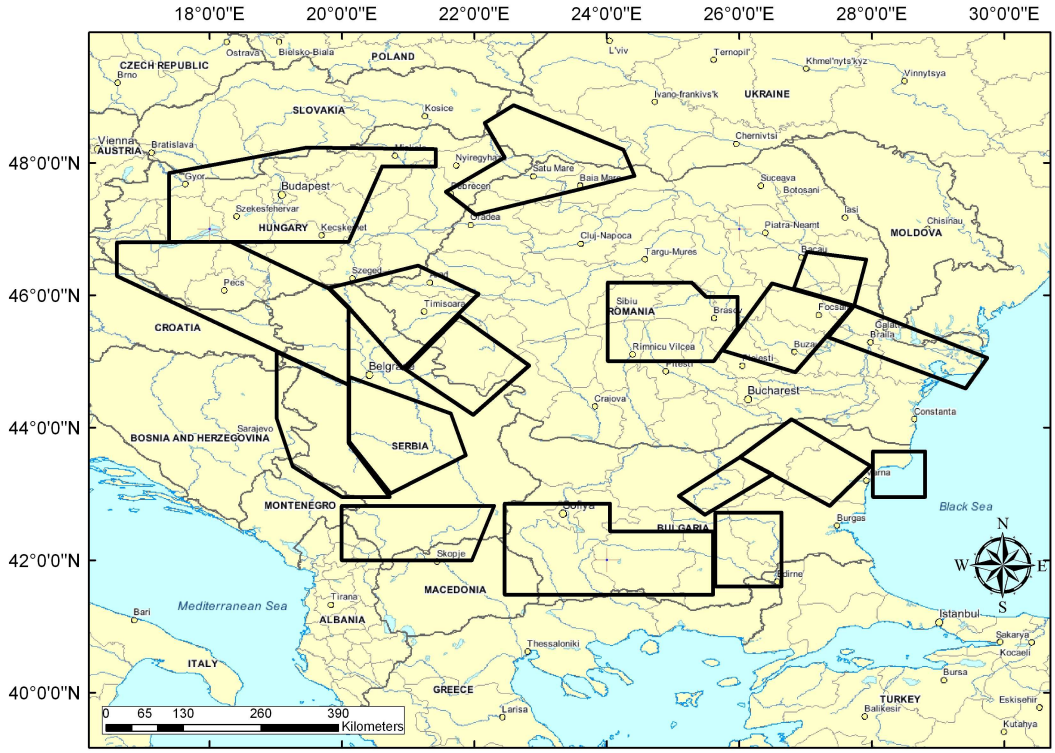

Figure 3.10: Seismic sources for the Balkan Region (Ardeleanu *et al.*, 2005)

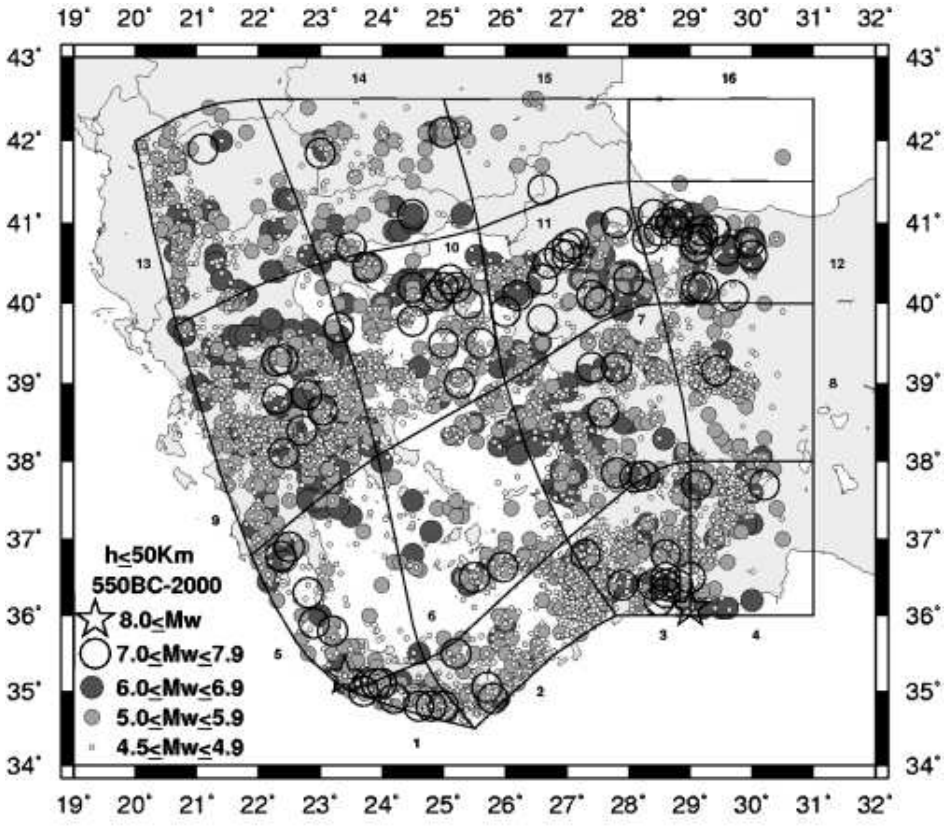

Figure 3.11: Geodetic deformation source model of Koravos *et al.* (2003b)

zones comparison is made between seismic and tectonic moment release, proceeding to define recurrence characteristics and maximum "credible" magnitudes for each zone. Whilst this approach is certainly of interest, the definition of these zones is not necessarily based on seismological properties, but deformation. Comparison could of course be made with such zones. Since it is already known that definition of zone boundaries can be a subjective process, often depending on the purpose for which the zones are defined, then treating them in the same manner as one would a source zone designed for hazard analysis is spurious, even if seismological properties can be defined.

3.5 Philosophy of Source Zonation

It is clear that despite the inherent subjectivity in seismic source zonation, the alternatives are not necessarily a significant improvement when it comes to replicating seismicity. Before proceeding to discuss the development of an original source model for the Aegean, there is one more issue to consider: the comparative role of knowledge and uncertainty. A key element that underpins seismic hazard analysis is the recognition that observed seismicity is not representative of all possible scenarios of future seismicity. In that sense, there is a case to be made for defining a model that allows for the possibility of extremely unlikely events, whilst at the same time retaining the higher probabilities for events similar to those more frequently observed. If this is so, one may choose to invoke Occam's razor: spatial distribution of seismicity is preferentially modelled by fewer zones, albeit allowing for large uncertainties in the zone parameters, than by large numbers of zones whose parameters are highly dependent on the small amount of observed seismicity.

3.6 Information used in the development of a new Aegean seismic source model

To produce a new source model for the Aegean region many sources of seismotectonic information need to be considered. These include observed seismicity, focal mechanism, geology and tectonic information. Much of this is found in published scientific literature and may address both region wide tectonics or seismicity and faulting within a small locality. The discussion of each sub-region within the Aegean is presented in Appendix A. Several region-wide data sets play an important role in influencing the delineation of seismic sources, which will be discussed here.

3.6.1 Observed Seismicity

The principal earthquake catalogue used in the development of these source zones is that of Burton *et al.* (2004a) and shown in Figure 3.12. This is extended to 2005 by addition of the NOA catalogue. The historical earthquake catalogue of Papazachos and Papazachou (1997) is also referred to for the inclusion of historical earthquakes. For the period 1900

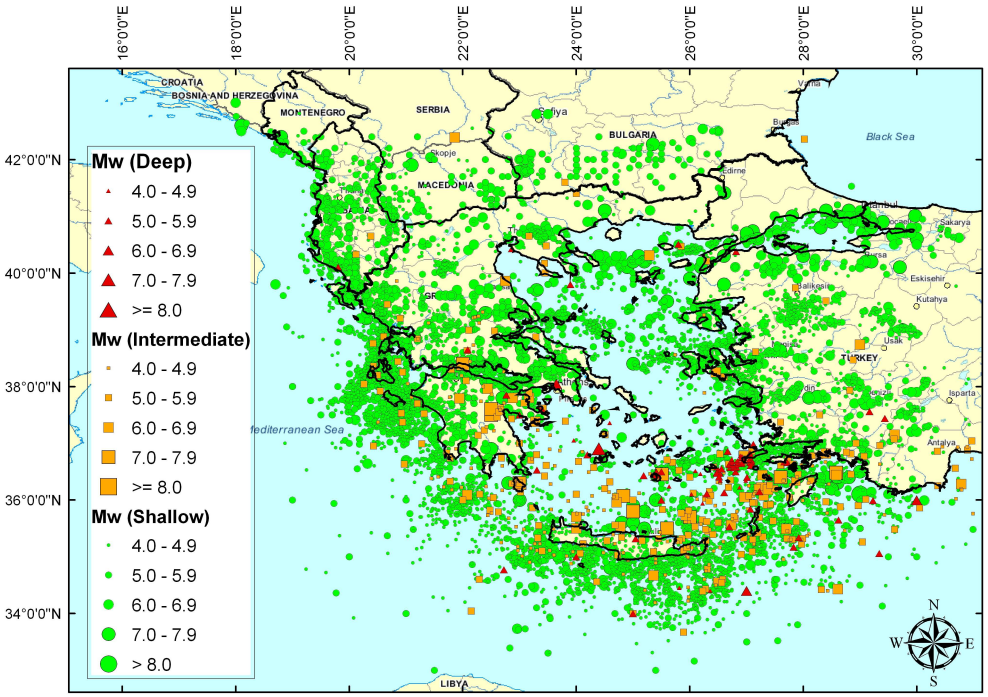

Figure 3.12: Observed Seismicity of the Aegean region for the period 1900 2005

- 2005, the earthquake catalogue may be considered complete above M_W 5.2. Refer to Chapter 2 for further analysis.

3.6.2 Focal Mechanisms

Several publications present fault plane solutions for parts of Greece and the Aegean. These are referred to where relevant, though perhaps the most extensive records can be found for the Adriatic region (Anderson and Jackson, 1987; Baker *et al.*, 1997; Louvari *et al.*, 2001), the Aegean (Papazachos *et al.*, 1991; Hatzfield *et al.*, 1999; Kiratzi, 2002; Kiratzi and Louvari, 2003), the Hellenic Arc (Benetatos *et al.*, 2004; Bohnhoff *et al.*, 2005) and the Sea of Marmara (Pinar *et al.*, 2003). Fault plane solutions from Kiratzi and Louvari (2003) are shown in Figure 3.13.

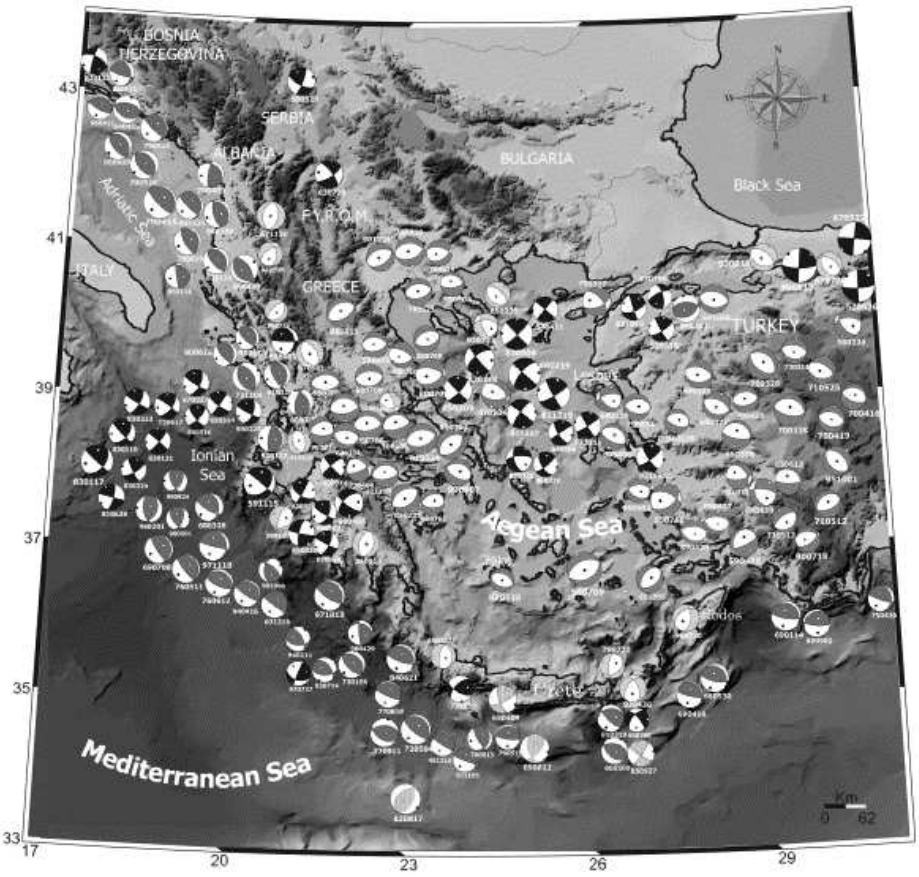

Figure 3.13: Focal mechanism of earthquakes in the Aegean region (Kiratzi and Louvari, 2003)

A database of Earthquake Mechanisms of the Mediterranean Area (EMMA) has been assembled by Vannucci *et al.* (2004) (Figure 3.14). This comprises focal mechanisms for the Mediterranean taken from the Harvard Global Centroid Moment Tensor (CMT) catalogue, the Euro-Mediterranean Regional CMT catalogue (RCMT) and the ETH Catalogue. The inclusion of earthquakes from regional catalogues expands the range of magnitudes from which useful mechanisms are determined. In some cases magnitudes may be as low as M_W 4.0.

3.6.3 Tectonic Models and Satellite based Geodetic Observations

The Aegean is arguably one of the most complex tectonic regions on the globe. Consequently, kinematic models of the tectonics have evolved over time. These interpretations often form the basis for the development of source models. There are many localities where these interpretations may conflict with each other, which the addition of more GPS data may assist in resolving. GPS velocity vectors are considered strongly in the development of tectonic block models, which then influence the zonation of sources for seismic hazard.

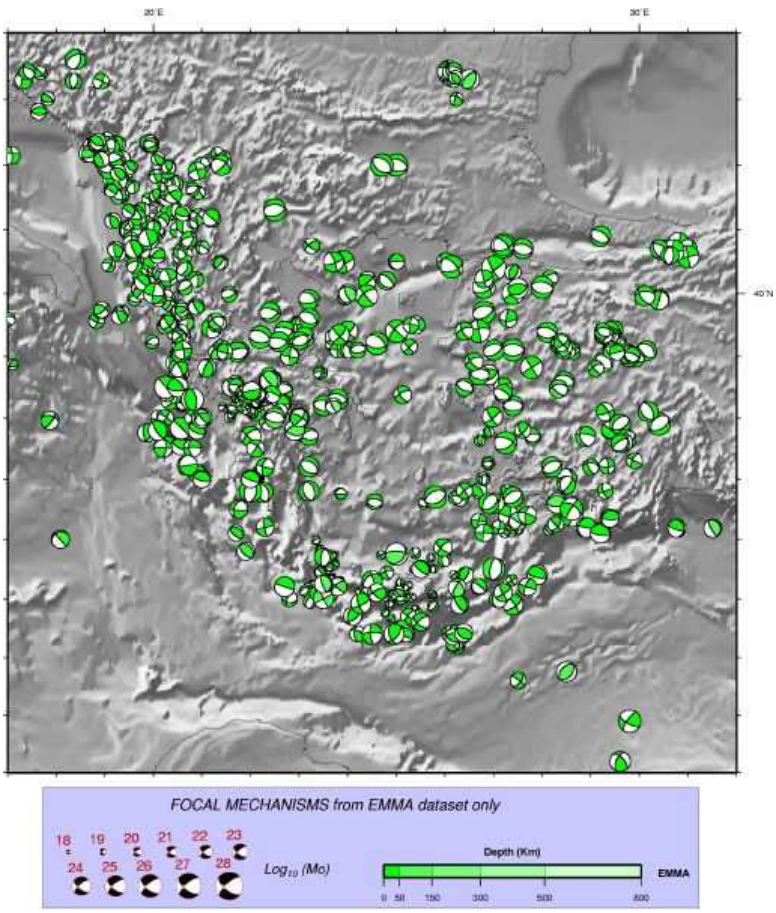

Figure 3.14: Focal mechanisms from the EMMA database (Vannucci *et al.*, 2004)

GPS studies that form a key basis of the development of the source model are those of McClusky *et al.* (2000) and Nyst and Thatcher (2004). The latter publication delivers a comparison of the various interpretations of Aegean geodynamics from various sources over the preceding 30 years. There is general agreement within these kinematic models of the role of the anti-clockwise rotation of the Anatolian plate. This generally accounts for the dextral strike-slip motion along the North Anatolian Fault across Northern Turkey and into the Sea of Marmara. Disagreement comes as to the extent to which the dextral strike slip motion of the NAF is apparent in the tectonics of central Greece and the North Aegean Sea. Most kinematic models require a block boundary that traverses northern and central Greece to connect the dextral slip of the North Aegean Sea to the plate margin of the Adriatic. As noted by Goldsworthy *et al.* (2002), this is not borne out by the observed seismicity.

Perhaps the most important geodetic study used in the design of a new seismic source model is that of Reilinger *et al.* (2006) (Figure 3.15). In addition to providing a GPS analysis of the Eastern Mediterranean and Middle East, a kinematic block model is also produced. Boundaries well constrained by observed seismicity and faulting are distinguished from those inferred from geodetic measurements. Such a detailed model assists greatly in the designation of seismic source zones. Stronger justification can be given for

delineating zones across regions of near uniform deformation. There are of course several block boundaries that are not in great agreement with the distribution of observed seismicity. These can be found mostly in the Northern Greece and South Balkan region, which is characterised by typically low rates of slip. Other margins ill-constrained by seismicity can be seen in central Crete and southwestern Turkey.

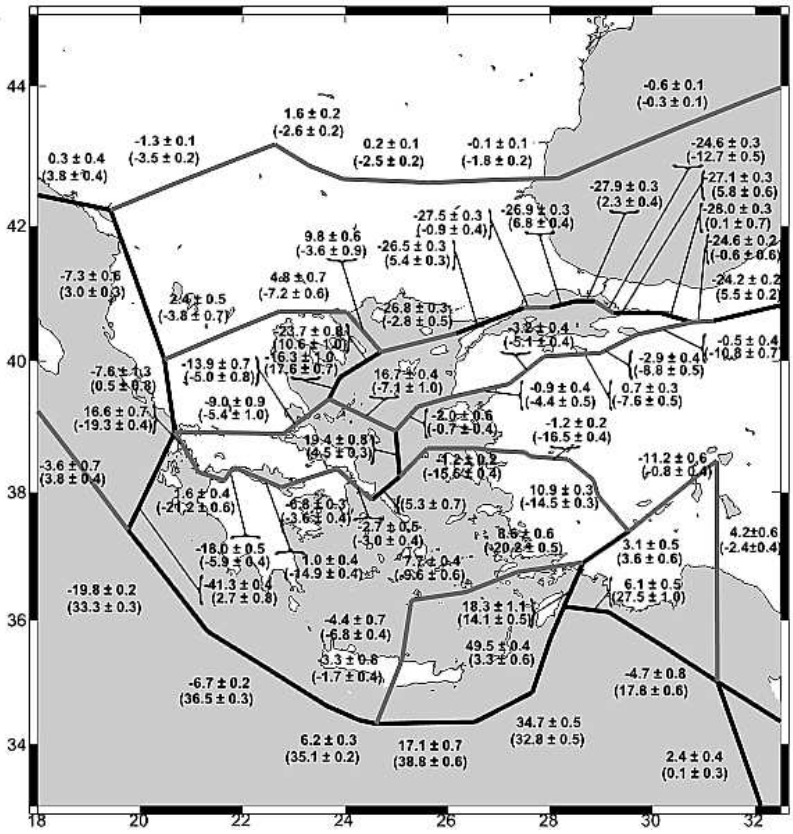

Figure 3.15: Geodynamic block model of Reilinger *et al.* (2006). Black lines indicate faults constrained from observed geology, grey lines are inferred faults. Numbers out of brackets indicate fault parallel slip (sinistral being positive), numbers in brackets indicate fault normal slip (compression being positive)

The new source model is presented in the next section. In the interests of transparency, a zone by zone discussion is presented in Appendix 1. Within the description is a comparison of the new zones with previous zonations in the region. This is also related to the observed seismotectonics and seismicity of each location.

3.7 The New Source Model (WT2006)

The new shallow seismic source model for the Aegean region is presented in Figure 3.16, with the intermediate-depth seismic source models shown in Figure 3.17. Seismicity parameters are given in Table 3.1. The coordinates for the zones can be found in the Appendix. The final number of shallow source zones defined is 15, with two seismically active areas left unzoned: Northern Greece/Southern Balkans, the South Aegean/Western Turkey.

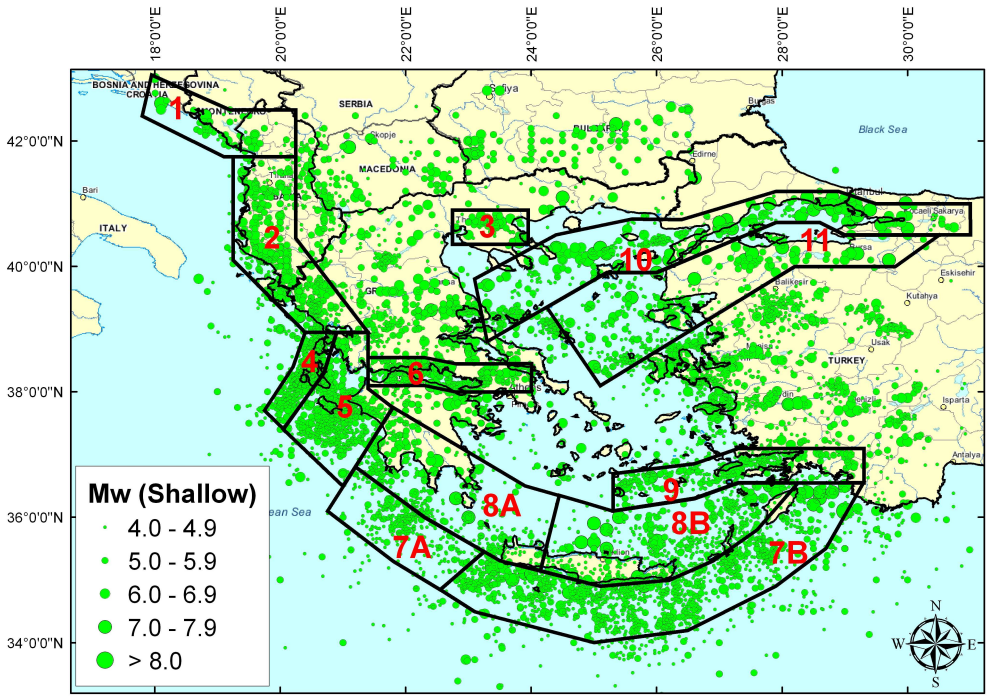

Figure 3.16: WT2006 shallow seismic source model superimposed over shallow Aegean seismicity

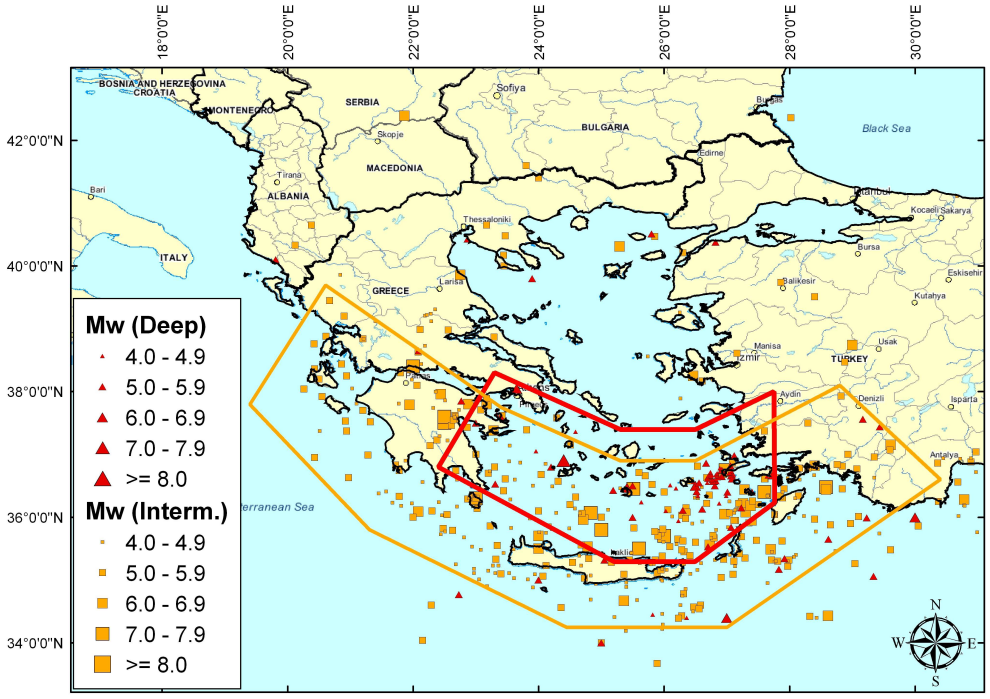

Figure 3.17: WT2006 intermediate-depth zones 60 to 100 km (orange) and 100 to 180 km (red), superimposed over intermediate-depth Aegean seismicity

Table 3.1: Summary of zones and seismic parameters for the WT2006 source model. LS = Least Squares, ML = Maximum Likelihood, M_C = Magnitude of Completeness (Maximum Curvature) and M_{MAX} Obs = Maximum earthquake (M_W) in historical catalogue

Zone	Name	b (LS)	σb (LS)	b (ML)	σb (ML)	M_{MAX} Obs	Mc
1	Montnegro	0.672	0.035	0.704	0.116	7.2	5.2
2	Albania	1.171	0.094	0.848	0.050	7.0	5.2
3	Thessaloniki	0.775	0.049	1.031	0.171	7.0	5.3
4	Cephalonia	0.907	0.046	1.004	0.079	7.4	5.2
5	Ionian	1.053	0.061	1.225	0.079	7.1	5.2
6	Corinth	0.864	0.064	0.909	0.073	7.0	5.3
7A	W.Hellenic	1.161	0.098	1.509	0.190	6.1	5.2
7B	E.Hellenic	0.740	0.019	0.768	0.038	8.3	4.8
8A	S.Greece	0.804	0.043	0.954	0.111	7.5	5.2
8B	E.Crete	0.600	0.036	0.673	0.102	8.2	5.5
9	CTEZ	0.762	0.062	1.021	0.133	7.2	5.2
10	NNAF	0.658	0.062	0.458	0.027	7.6	5.3
11	SNAF	0.695	0.062	0.578	0.096	7.4	5.3
DZ 1	DZ1	0.613	0.024	1.012	0.123	7.8	5.5
DZ 2	DZ1	0.870	0.054	1.169	0.242	7.0	5.3

When comparing this source model to the two most recent zonations of PP2000 and PZ1990 (Figures 3.18 and 3.19 respectively), the most obvious contrast is in the number of zones. The WT2006 model has only 15 zones of shallow seismicity (and two of intermediate seismicity), in contrast to the 36 zones of shallow seismicity (six intermediate depth) found in PZ1990, and 67 zones of shallow seismicity (seven intermediate depth) found in PP2000. Arguments for this have been presented within the description of the model, but they ultimately stem from two philosophies. The first is that one should only assign zones where there is an abundant amount of information about the coherence, nature and style of faulting. Where information regarding the physical nature of the faults is lacking and only the seismicity is used, one cannot be certain that the seismicity observed is stationary or an artefact of a short catalogue.

The second philosophy is that where seismicity parameters of neighbouring zones differ by a small amount (e.g. less than the standard error of the parameters), there is little ground to justify separating these zones. It should be noted that calculations of the seismicity parameters for each zone in Papazachos (1990) and Papaioannou and Papazachos (2000) are not well analysed in their respective references. In the former, b-values are simply assigned to the zones with seismicity rates dependent on rates observed, in the latter the b-values are smoothed across neighbouring zones. In both cases this results in a source model with neighbouring zones having near identical parameters, only distinguishable by the observed rates of seismicity. As the zones are in close proximity, it is debatable whether it is reasonable to divide them on the basis of observed rate, which may not take into consideration time-variant properties of earthquakes in the region.

Despite the fact that the WT2006 model is actively trying to implement a simpler zona-

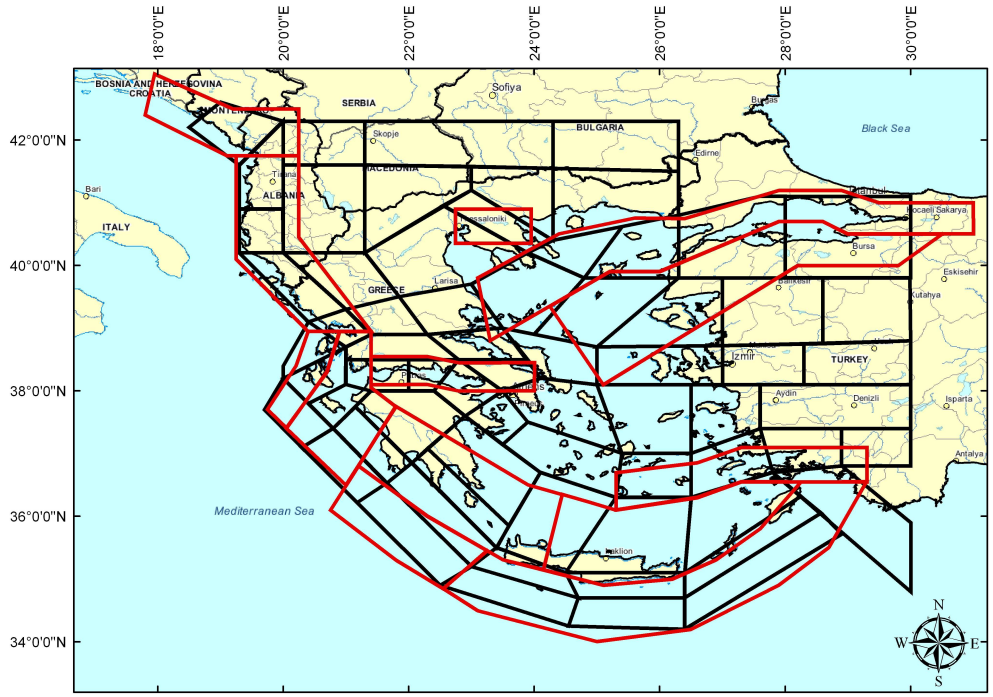

Figure 3.18: WT2006 source model (red) superimposed on PP2000 source model (black)

tion, there are many similarities between this and the previous models. Comparison with the PP2000 model (Figure 3.18) and PZ1990 model (Figure 3.19) clearly show that along the Hellenic Arc and Adriatic coastline, the Gulf of Corinth, and southwestern Turkey, single source models in the WT2006 model are a reasonable approximation to an amalgamation of neighbouring zones. The most notable differences lie in the NAF region, Thessaloniki and the Gulf of Corinth. Given the second philosophy of this source zonation this is to be entirely expected. It may not be unreasonable to suggest that if one was to create yet another source model by simply amalgamating certain zones in the PP2000 model, the differences between this and the WT2006 may be less significant. The good correlation of some zone boundaries between models suggests a degree of robustness in these zonations.

Where the WT2006 model may be particularly controversial is in the delineation of "cluster" zones. To all intents and purpose, this is simply the same assumption made in the PZ1990 for locations where source zones were not defined. These were classed as "background". This approach was reconsidered following the Kozani-Grevena (1995) earthquake and Athens (1999) earthquake, whose epicentres were outside the source zones shown. One thing that must be made clear, even if it is an issue of semantics, is that in the WT2006 model the term "background" is nowhere to be found.

Typically in PSHA, the term "background" can refer to one of a variety of interpretations. The first can simply mean that in this region, no seismicity of significance for the hazard analysis will be observed. A second interpretation, and one used frequently in

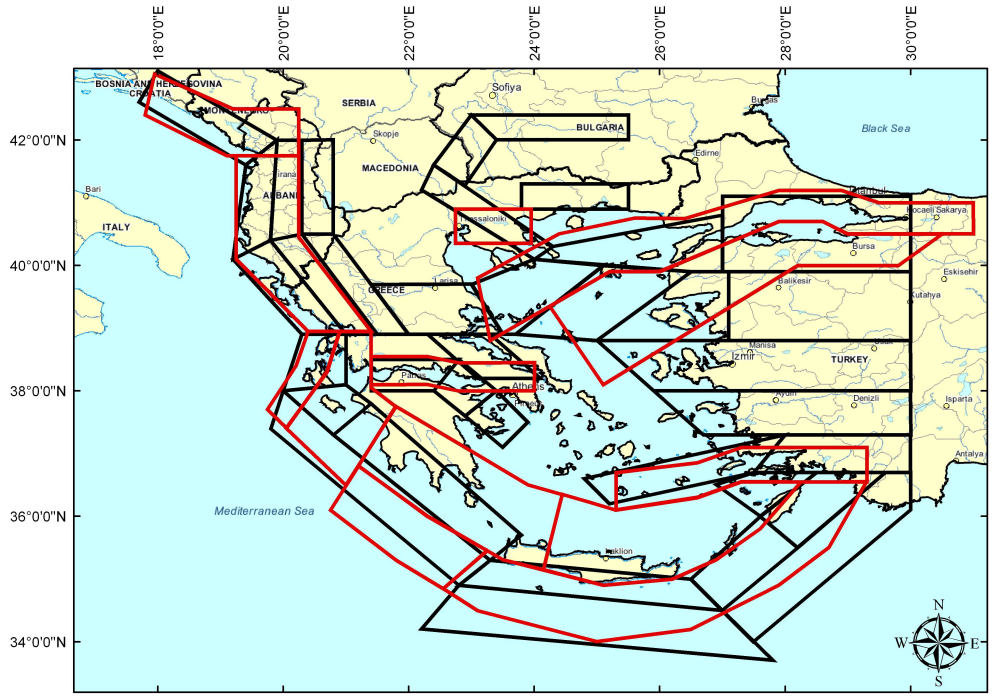

Figure 3.19: WT2006 source model (red) superimposed on PZ1990 source model (black)

hazard maps, is that background simply refers to another large uniform zone, within which earthquakes may have a very large recurrence interval but otherwise display seismic properties characteristic of any uniform zone. Neither of these is an appropriate description of the non-zoned regions presented here.

These are simply defined as regions where there lacks sufficient evidence to suggest that observed seismicity is not indicative of future seismicity. These zones are active and they must display similar properties in terms of focal mechanism or broad tectonic regime. This much is true for the "cluster" zones shown here. Both the Northern Greece and South Aegean regions display extensional faulting, with the primary axes of tension in a North-South direction. Equally, both regions are characterised by lower strain rates than for other regions of the Aegean, which results in longer return periods for the largest earthquakes. In both regions strong seismic activity has been observed and it may be possible to attribute this to existing faults. However, the complexity of the fault systems means that it is not possible to assert with the same degree of certainty how these faults will behave in future. This means that the only information available is the observed seismicity.

It is not just coincidence that the "cluster" zones are all within intra-plate regions. Within these regions some faults may be defined, but the low seismic activity makes it hard to determine recurrence parameters. This makes the two non-zoned regions of the Aegean (excluding the outer arc, whose seismicity is too sparse) ideal candidates for application of zone-free methodologies. This is, to some extent, a hybrid approach whereby spatial

uniformity is assumed for areas with well-known active faults, whilst at the same time spatial clustering is captured where fault systems are not well defined. The question remains as to what methodologies can be applied, a discussion of which has already been given in section 3.3.

The seismic hazard approach that will be the focus of future work using this source model is the Monte Carlo approach. This can incorporate two types of source, the first being a zoned model, the second being the observed clusters. Where seismic sources have been defined in this model, the Monte Carlo approach will create synthetic catalogues of earthquakes whose epicentres are sampled at random from the uniform zone. Within these catalogues the non-zoned seismicity is also simulated by randomly re-sampling (with a Gaussian error and replacement) the hypocentres of the observed seismicity to produce the synthetic earthquakes. This will produce synthetic catalogues whose spatial properties should be similar to the observed seismicity, whilst still allowing for a large amount of variability. More work will follow on this particular topic.

Although the zone-free methodologies are attractive, they have a particular drawback in that it is not possible to ascribe Gutenberg-Richter parameters to the seismicity. One has to assume implicitly that the observed magnitude distribution of the cluster zone is sufficiently representative of the true magnitude distribution over the appropriate timescales. What may be required is a way of identifying regions of seismicity with similar properties, but whose spatial distribution cannot be considered uniform.

The material presented in Appendix A provides a very detailed analysis of the seismicity and tectonics of the Aegean. This serves as a basis for the development of a new model, where the decisions behind the delineation of particular zones are transparent. It remains a subjective process, relying on a particular approach or philosophy to zonation that had been proposed at the beginning of the process. The aim of this work has been to present a rigorous discussion of the decisions upon which this zonation is based. It may be perceived that within this work the seismotectonic, geodetic and geological information have been analysed and presented, yet arbitrary decisions have still been made regarding the location of zones based on personal judgement. The question remains as to how this model can be made more robust.

Where a seismic hazard analysis demands a new source model it is common practice to assemble a panel of experts to each undertake a similar analysis. This would be welcomed with regard to the seismic source model for Greece presented here. For the Aegean there is an abundance of published literature pertaining to the geology, seismotectonics and geodynamics of the region, to say nothing of one of the most extensive historical catalogues available. Elsewhere this may not be true, even in seismically active areas. Where there is a poorer body of information upon which to develop models of the seismic sources, zonation on the basis of observed seismicity becomes an even more subjective process.

Some of the common problems when zoning based on observed seismicity lie in the characterisation of the seismic source. Earthquake catalogues generally describe earthquakes in terms of the hypocentral location, time and magnitude. This results in a set of point sources, often leaving little information regarding the strike and dip of the faults upon which the earthquake occurred. This may be supplemented by observations of the fault using InSAR or ground-based geodetic measurement, made easier if the fault scarp is visible. The focal mechanisms of earthquakes provide important physical information, but it may not always be possible to distinguish between the fault plane and auxiliary plane. These are approximations of the nature of the seismic source, and the quality of the data may vary considerably even within a small study region.

In the absence of geological, seismotectonic and geodetic knowledge in a region, the problem of how to define seismic source zones still remains. Clearly the earthquake catalogue provides the greatest source of information, but delineation of zones on the basis of the observed seismicity alone may be highly subjective. In the next chapter a K-means cluster analysis approach is introduced as a means of objective delineation of seismic source models. This will illustrate how physical properties of earthquakes can be incorporated into the analysis.

Chapter 4

Partitioning Seismic zones using Observed Seismicity: A K-means Cluster Analysis Approach

4.1 Limitations of the Source Zonation Process

A new seismic source model has been presented in the previous chapter, alongside discussion of many of the issues surrounding seismic source modelling. Comparison of the most widely used seismic source models for the Aegean was also made. It should be apparent that whilst the source models may reflect the state of knowledge of seismotectonics within the Aegean, they are also subjective. Differences between particular models emerged as a result of changes in observed seismicity, the context for which the source model was being applied, and the philosophy of the zonation. These variations also assume that the information being input into the zonation process is consistent.

The subjectivity inherent in the process of seismic source zonation is an important shortcoming of seismic hazard analysis, but not one that is adequately addressed when analysing epistemic uncertainty. It is common for several contrasting source models to be implemented in a logic tree analysis of uncertainty in seismic hazard. The impact this has on the total epistemic uncertainty is not always readily addressed. Often, weightings of different source models are ascribed by expert judgement. In low seismicity regions, source models can conflict substantially. As noted by Beauval *et al.* (2006b), for regions where faulting is not well-characterised, "different experts often provide very different maps that characterise somewhat different zonation schemes, based on the differing interpretation of the meagre data that exist".

The Aegean is an interesting region for considering the nature of seismic source zona-

tion. This is because of the large disparity in seismic behaviour, and the contrast between hazard from seismically actively plate margins and lower seismicity intra-plate regions. Whereas faults in other plate margin regions may be well constrained by geophysical investigation or by visible rupture, for many areas in the Aegean the most active sources are offshore (Danciu and Tselentis, 2007). This includes the Hellenic arc, the Ionian Islands and the North Anatolian Fault in the Sea of Marmara. Although many seismically active faults can be found onshore in the Aegean region, a region-wide fault database has not yet been developed. This is due to differing interpretations of the observed faulting, and inconsistencies in fault delineation across political boundaries. For site-specific engineering applications it is possible to identify seismic sources critical to hazard in the Aegean, but not for seismic hazard mapping applications. Nowhere is this more evident than in the Northern Greece and Southern Balkan region of the Aegean. Only the hypothesised sources of major historical earthquakes in this region give any indication as to the potential location of future seismicity.

The need for a model of areal seismic sources across the Aegean has been discussed in the previous chapter. Given the shortcomings and subjectivity of the existing approaches to seismic source zonation, the question arises as to how this can be improved so as to enhance the accuracy of future seismic hazard analyses in the Aegean.

4.2 Expressing seismic source delineation as a partitioning problem

The objective of seismic source zonation had been to divide the source region into a set of zones, each zone containing uniformly distributed seismicity. To do so, it is necessary to analyse variation in a spatially distributed set of observed seismotectonic data. This could include focal mechanism, stress/strain vectors, geology and, perhaps most importantly for seismic hazard analysis, observed seismicity. Each datum within the data sets being considered may contain a large number of parameters. For example a single earthquake focal mechanism may contain more than 14 parameters (Longitude, Latitude, Depth, Moment, Nodal Plane 1 (strike, dip and rake), Nodal Plane 2 (strike, dip and rake), pressure axes (azimuth, dip), tension axes (azimuth, dip), stress drop etc.). With such a large number of inter-dependent parameters it is difficult to identify similarity by an algorithmic process. When integrating many different data sets, it becomes near impossible to do so.

There are many sources of spatially distributed information from which to describe the seismotectonics of a region. Each source has a varying degree of completeness over a large area. For example, evidence of stress and strain from GPS data is confined to the area covered by the existing GPS network. Across the Aegean there are only a few hundred GPS stations, most of these are in continental Greece and Turkey, with only a small number found on islands in the Aegean Sea. Since many of the most active sources are

offshore, a large area of the Aegean is poorly characterised. Equally, the degree of completeness, and accuracy of focal mechanism data, varies enormously depending on the observed seismicity and the quality of the seismic network.

The simplest set of observed data is the seismicity catalogue. This generally characterises an earthquake in terms of three dimensional location, time and magnitude; a sum total of five dimensions. Spatial and temporal variation in the completeness of observed seismicity is something that can be analysed numerically (see Chapter 2). Hence, it is possible to analyse observed seismicity as a coherent data set in a manner that is not necessarily true for other seismotectonic information.

Earthquake catalogues define the seismic source in terms of a point source: the hypocentre. This is generally the origin of the earthquake energy release. However, for larger earthquakes a point-source origin of energy release is not necessarily appropriate for defining the origin of the earthquake. For larger events ruptures of the order of tens to hundreds of kilometres are common, in which case a point-source is not appropriate. This will be considered in more detail further on.

4.3 K-means Cluster Analysis

K-means cluster analysis (Hartigan, 1975; Hartigan and Wong, 1979; Jain *et al.*, 1999) is an example of a hard partitioning algorithm. A set of N data ($x_1,\ x_2\ \ldots\ x_N$) in d dimensions is partitioned into K clusters, where each element in the data set is allocated entirely to a particular cluster. It is an iterative process whereby the data are initially partitioned, the mean position of each group calculated, and then the data partitioned again by allocating each datum to its nearest mean cluster position. The procedure terminates when no datum changes cluster or when the number of iterations reaches a pre-defined maximum (usually 100 iterations, as is the case here). The algorithm is described by the flowchart in Figure 4.1. In this research the K-means algorithm used is a modified version of the K-means code provided by Nabney (2002), implemented in Matlab 7.1.

4.3.1 Distance Metric

In most applications of K-means clustering the preferred distance metric is the Euclidean square distance (Hartigan and Wong, 1979). For many practical applications of K-means cluster analysis the Euclidean-square metric is still the most common (Jain *et al.*, 1999). This is a special case of the Minkowski distance metric, with p = 2:

$$Distance_{p-norm} = \left(\sum_{i=1}^{n} (x_i - y_i)^p \right)^{1/p} \tag{4.1}$$

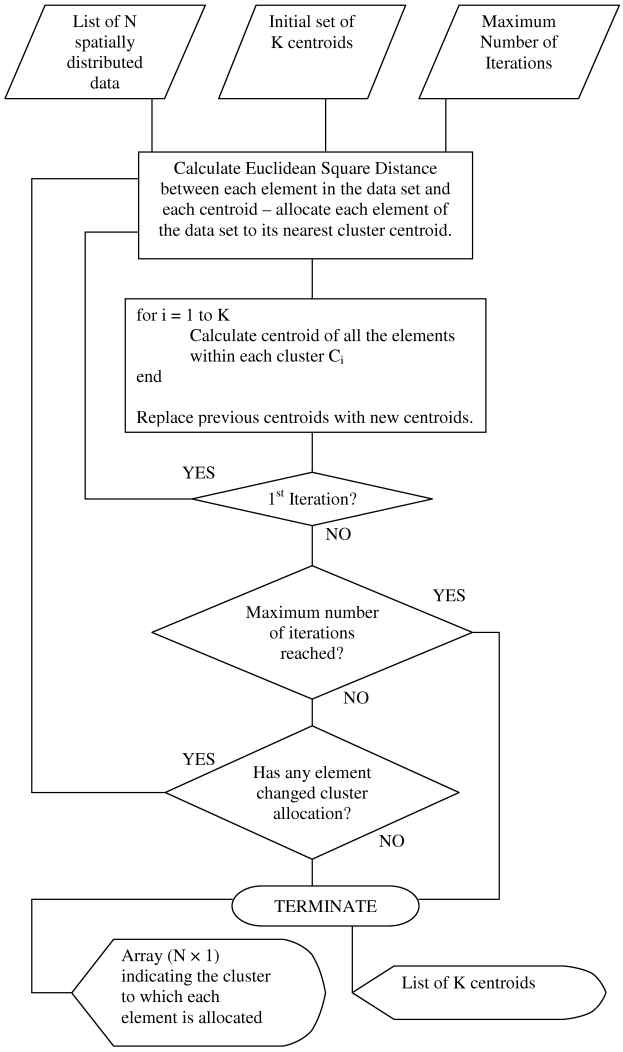

Figure 4.1: Flowchart of the K-means algorithm

$$d(\mathbf{x},\mathbf{y}) = Distance_{2-norm} = \left(\sum_{i=1}^{n}(x_i - y_i)^2\right)^{1/2} \tag{4.2}$$

The Euclidean distance metric is arguably the most intuitive of the Minkowski metrics to use in this application of cluster analysis. This is simply due to the use of Euclidean space, in which basic Pythagorean geometry is applied. Theoretically there is no upper limit to the number of dimensions of data that can be clustered using K-means, though only 2- and 3-dimensional spatially distributed data shall be considered here. Since the objective of this application is to partition a set of earthquakes distributed across a fixed crustal volume, more complicated distance metrics may not be appropriate. They also offer little compensation for the additional computation required.

When using Euclidean-square distance as a metric there is a risk that the largest scaled features will dominate the clustering process. This is usually overcome by normalisation of the data or by implementing weighting schemes to compensate for dominance of one

particular dimension. In the study of seismicity, an acceptable partition or set of clusters should provide viable seismic zones for seismic hazard quantification and analysis.

4.3.2 Quantifying the partition quality

Assessment of the "quality" of a partition is an important consideration in cluster analysis. The most common measure of cluster quality (for known K) is the total within-cluster sum of squares (TWCSS), also referred to as squared error or clustering error (Likas *et al.*, 2003). The TWCSS is defined as:

$$TWCSS = \sum_{i=1}^{N} \sum_{k=1}^{K} I\left(x_i \in C_k\right) \left\| x_i - m_k \right\| \tag{4.3}$$

Where m_k is the mean of cluster C_k, and I(X) = 1 if statement X is true, 0 otherwise. An alternative to this particular measure is the pooled within-cluster sum of square distances (WK). This metric is defined by Tibshirani *et al.* (2001) as:

$$D_k = \sum_{i,i' \in C_k} d_{ii'} \tag{4.4}$$

$$WK = \sum_{k=1}^{K} \frac{1}{2n_k} D_k \tag{4.5}$$

where n_k is the number of elements within cluster C_k, and $d_{ii'}$ is the Euclidean-square distance between point x_i and $x_{i'}$. For a cluster of n_k elements, TWCSS and WK can be visualised in the manner of Figure 4.2. Both TWCSS and WK provide information about the overall compactness of clusters, with compact clusters producing lower values for both indices. Neither index explicitly characterises the separability of the clusters in the partition. This requires another index, known as between-cluster sum of squares (B):

$$B\left(K\right) = \sum_{i} \left|C_i\right| \left(m - m_i\right)^2 \tag{4.6}$$

where C_i defines the number of data in cluster i, m_i the centroid of cluster i and m the centroid of the entire data set. There are also many other clustering indices, each designed to assess particular qualities of the partitions such as compactness and separability. One of the more commonly used indices is the silhouette index (Kaufman and Rousseeuw, 1990):

$$s\left(i\right) = \frac{b\left(i\right) - a\left(i\right)}{\max\left\{a\left(i\right), b\left(i\right)\right\}} \tag{4.7}$$

a(i) is the mean distance between a point x_i and all the other points within the same cluster, and b(i) the mean distance between point x_i and all the points in the next nearest cluster. For a specific partition the silhouette index is the mean s for the data set.

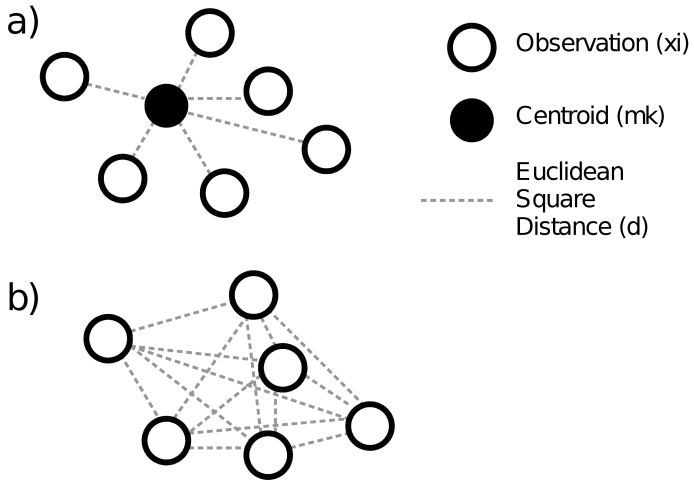

Figure 4.2: Schematic of within-cluster indices. a) TWCSS, b) WK

4.4 Identifying the Optimum Partition

Though the K-means algorithm is popular in several branches of science (e.g. image processing, pattern recognition and genetics), it suffers from two major problems that, despite the development of research into the procedure, remain unresolved. The first problem is that of determining the optimum set of initial centroids, given a specified K. The second problem is that for many applications, the most appropriate number of clusters (K) is not known *a priori*. Since most hard and fuzzy partitioning algorithms require K as an input, identification of the optimum number of clusters in a data set has become a key challenge in the field of cluster analysis.

A crucial difficulty in the development of methodologies for finding the optimum partition for a set of data is that the identification of the optimum K and the optimum seeds are not independent. Many procedures for identifying the latter assume that K is known *a priori*; those for the former assume that the optimum partition can be found for each specified K. In reality these problems are not independent. Application of algorithms to identify both the optimum seeds and optimum K require a large amount of computation. It is often the case that the global optimum partition for a data set, with a specified K, is not recognised. In such a case, it is a near-global optimum partition used; one that produces a lower TWCSS than a single implementation of K-means, but higher than the global optimum partition. For the practical applications of data partitioning, the use of a near-global optimum as opposed to a global optimum partition may not be significant. Where the optimum K may be determined using indices that are a function of TWCSS, it is possible that the results will not be robust to the choice of initial seeds. This means that the results may not be exactly replicable.

4.4.1 Initial Cluster Centres

It has been established that the aim of cluster analysis is to partition a set of data into K clusters. It is also true that K-means is an iterative algorithm that will produce the same result providing the inputs are kept the same (i.e. a local optimum, not global). As such, it is necessary to identify the initial seeds that will result in the global minimum TWCSS. Many of the approaches to overcome this employ stochastic techniques to identify the optimum set of initial cluster centres (Forgy, 1965; Bradley and Fayyad, 1998; Peña *et al.*, 1999). Alternatives to stochastic techniques tend to require either exhaustive searches for the optimum initial seeds (Likas *et al.*, 2003), which can be computationally impractical even on high performance computers, or a change in the definition of the cluster centre (Zhang *et al.*, 1999).

The simplest approach to identify the optimum initial clusters is an ensemble analysis (Peña *et al.*, 1999; Kuncheva and Vetrov, 2006). The algorithm is repeated a large number of times, with different initial seeds (randomly selected) on each trial. The partition that produces the smallest TWCSS is the optimum. As is common in stochastic search procedures, this optimum may still only be a local rather than global optimum. However, if the size of the ensemble is large enough, it may be sufficiently close to the global optimum for practical purposes. In this work ensembles with no fewer than 100 implementations of the K-means algorithm are used. Tests on synthetic data indicate that this is generally sufficient to determine a good near-optimum partition, and, when considering data sets of fewer than 10,000 points, appears to be robust.

Krishna and Murty (1999) and Lu *et al.* (2004) extended the stochastic approach to a Genetic K-means algorithm, which converges toward the global optimum partition using evolutionary computing algorithms. This tests a large number of random solutions and selects the better fitting ones as a basis for a new population of solutions. It is suggested that cluster analysis is a particularly pertinent application of genetic algorithms, as a partitioned data set can be readily encoded into a chromosomal format (i.e. a string of numbers with each element corresponding to the cluster allocation for each datum). It needs to be recognised that in a purely genetic approach the K-means algorithm is not run to iterative convergence. Krishna and Murty (1999) use the genetic algorithm to converge towards a global minimum of TWCSS, without using any additional heuristics. Lu *et al.* (2004) instead use a single iteration of K-means as a heuristic to accelerate convergence and improve performance. This has the impact of exploring more of the data space, resulting in a higher confidence that the partition is a global optimum. These genetic approaches to cluster analysis are still in their infancy and it is hoped that with improving processing power that they may become more widely used.

When the ensemble K-means method is implemented, consideration must be given as to the method of random initialisation. Comparison of several methods is given by Peña *et al.* (1999), and the stability of the algorithm with respect to the initialisation method

analysed by Kuncheva and Vetrov (2006). The two most common approaches to initialisation, and ones commonly found in commercial K-means software packages, are Forgy (1965) initialisation and Random Partition (RP) initialisation. Forgy (1965) initialisation requires that K initial cluster centres be selected by random sampling (without replacement) the observed data set. RP initialisation requires that each datum is allocated to a cluster at random. To avoid bias in cluster sizes at the initialisation stage N/K data points are allocated to each cluster. Where K is not a factor of N, the remainder of points are randomly allocated to clusters. Although both methods are stochastic, they can differ in the manner in which data are then subsequently partitioned. For most data sets, when using RP initialisation the centroids will converge toward the middle of the data space after the first K-means iteration. When Forgy initialisation is used there is an increased likelihood that outlying data points are used as initial centres. When this happens it is not uncommon for the K-means algorithm to produce empty clusters. It is found that RP initialisation tends to result in more stable partitions (Bradley and Fayyad, 1998; Peña *et al.*, 1999).

4.4.2 Identifying Optimum K

Many clustering algorithms require specification of the number of clusters *a priori*. In many applications, including this one, it is not obvious how many clusters a data set contains. This is a classic issue in cluster analysis and many different approaches have been suggested in attempts to solve it (Tibshirani *et al.*, 2001; Feng and Hamerly, 2006). These approaches can be subdivided into three different classes:

1. Validity indices that do not correlate with K (Calinski and Harabasz, 1974; Krzanowski and Lai, 1988; Kaufman and Rousseeuw, 1990; Tibshirani *et al.*, 2001).

   Validity indices that depend on within-cluster compactness have been considered previously, i.e. TWCSS and pooled within-cluster sum of square distance. These indices are not suitable for identifying optimum K, as they are inversely proportional to increasing K. This is because these indices are function of the within-cluster sum of square distances, which decrease as the number of clusters increases and the average number of elements within each cluster decreases. Suitable indices for identifying the optimum K should not correlate with K. A description of some candidate indices can be found in Milligan and Cooper (1985) and Tibshirani *et al.* (2001). Those indices tested in this research are discussed in more detail in section 4.5.4.

2. Cluster splitting depending on a failure criterion Hamerly and Elkan (2003); Welling and Kurihara (2006); Feng and Hamerly (2006).

   Here, statistical data tests are applied to the data to test whether it can be considered as conforming to a particular distribution (usually Gaussian, as is the case for K-means). Starting with the null case (K = 1, all the data are within one cluster and should not be partitioned), the cluster is tested to see if it is distributed in a

Gaussian fashion. If it passes the statistical test at a specified level of confidence the cluster remains, otherwise the cluster is split and the test applied to each of the two new clusters. This process will continue until all clusters pass the test (hence the optimum K) or until a specified K_{MAX} is reached. These methods can be quite powerful if the clusters are reasonably well separated, but often over fit the data if clusters are poorly separated. Where clusters may appear to overlap, the optimum K estimated is strongly dependent on the statistical test used and the probability level at which the test is passed (Feng and Hamerly, 2006). They also tend to assume that clusters are ostensibly Gaussian in nature, which is not always the case.

3. Genetic K-clustering methods with variable K (Sheng and Liu, 2006)

   These are useful variations of the genetic K-means algorithm (Krishna and Murty, 1999; Lu *et al.*, 2004), using K-medoid clustering as a heuristic. K-medoid clustering is a variant of the K-means algorithm where the nearest element to the cluster centroid becomes the cluster centre, as opposed to the centroid itself. By including K as a variable, the genetic K-medoids algorithm should converge to a global optimum partition for K in the specified range. Much like the genetic K-means algorithm of Lu *et al.* (2004), K-medoids is used as a heuristic to speed convergence. These genetic cluster analysis algorithms could prove useful in future, but are currently highly computationally intensive. They are also still early development and it remains to be seen how well they perform on different real-world data sets.

Ultimately, the choice of K may still be an expert decision based on the output from different indices and methods. In this research, flexibility, computational efficiency and feasibility of coding on non-parallel computing platforms are crucial considerations. As such, validity indices are used here as a means of identifying the optimum K. Several indices that performed well when identifying the correct number of clusters in a synthetic data set (well-separated clusters with known K) were tested. A description of these indices can be found in Tibshirani *et al.* (2001) and some discussion is given further on.

4.4.3 Illegal Partitions

Illegal partitions are defined as partitions for a specified K that will, after iteration, produce at least one empty cluster. This is a common hazard of cluster analysis and the likelihood of illegal partitions increases as $K \rightarrow N$. In the earliest K-means algorithms (Hartigan and Wong, 1979), an illegal partition would produce an error, which in ensemble analysis would be flagged and excluded from further consideration. The frequency of illegal partitions for high values of K is problematic, as it reduces the number of valid partitions from which the global optimum is determined. Although one could simply proceed with iteration, albeit with K-1 clusters; when identifying the optimum K this is undesirable. Instead a "singleton" procedure is invoked. This procedure requires that, on production of an empty cluster, the point furthest from its allocated cluster centre is

then selected as a new cluster, upon which the iteration proceeds. This has the impact of slowing down convergence of the K-means algorithm, but this is a satisfactory cost if it allows for more stable partitions at higher K.

It is entirely possible for the output of the K-means algorithm to produce singleton clusters. When this occurs it may be necessary for the user to decide whether to accept the singleton cluster or whether to re-merge the singleton with the nearest cluster. This will reduce the value of K and is therefore inappropriate when identifying the optimum K. Several indices of partition quality are a function of $min\left(x_i, m_k\right)_{x_i \in C_k}$, which may produce errors in the index if singleton clusters occur, and must be trapped in an appropriate manner.

4.4.4 Earthquake and Rupture Catalogues

The earthquake catalogue used in this analysis is the 20th century catalogue of Burton *et al.* (2004a), introduced in chapter 2. Only shallow earthquakes, with focal depths less than 60 km, are considered here. The completeness magnitude M_C for the 1900 - 1999 used for this analysis is M_W 5.2. Comparison needs to be made between the original catalogues and those for which non-Poissonian events have been removed. The preferred algorithm for removal of aftershocks is that of Reasenberg (1985).

Also considered in this analysis is a catalogue of known ruptures for historical earthquakes in the Aegean (Papazachos *et al.*, 1999). The ruptures in this catalogue are expressed in terms of their source parameters, including strike, dip and rake. These earthquakes are all assumed to be shallow events whose rupture planes have been distinguished using macroseismic intensities, surface observations or aftershock distribution. The catalogue found in Papazachos *et al.* (1999) contains 150 events up to 1995. This catalogue is cut-off at M_W 6.0. Using the atlas of isoseismals for Greece (Papazachos *et al.*, 1997), a further 64 fault ruptures are identified for earthquakes with magnitudes in the range $5.5 \leq M_W \leq 6.0$. For earthquakes in this magnitude range only the rupture strikes can be determined, dip is assumed to be 90°. For all but a few of the events in this catalogue the exact shapes of these faults remain largely unknown. An approximation to a line source is therefore made. The orientation of the lines is indicated by strike and the lengths calculated using the empirical fault scaling relation of Wells and Coppersmith (1994). The equation used is the 50^{th} percentile of the regression of the common logarithm of subsurface rupture length (RLD) against M_W for earthquakes of all fault types:

$$\log\left(RLD\right) = \left(0.59 M_W - 2.44\right) \pm 0.16 \tag{4.8}$$

This catalogue is supplemented further with an additional eight ruptures from significant earthquakes in the period 1996 to 2006: Dodecanese Islands, 20 July 1996 (M_W 6.2); Southern Greece, 13 October 1997 (M_W 6.4); Southern Ionian Islands, 18 Novem-

ber 1997 (M_W 6.6); Izmit, 17 August 1999 (M_W 7.6); Athens, 7 September 1999 (M_W 5.9); Skyros, 26 July 2001 (M_W 6.4); Lefkada, 14 August 2003 (M_W 6.2) and Kythera, 8 January 2006 (M_W 6.7). The fault dimensions (i.e. strike, dip, rake and subsurface rupture length) have been determined for four of these events: Izmit (Barka, 1999; Reilinger *et al.*, 2000; Wright *et al.*, 2001; Barka *et al.*, 2002), Athens (Papadopoulos *et al.*, 2000; Pavlides *et al.*, 2002), Skyros (Papadopoulos *et al.*, 2002) and Lefkada (Papadopoulos *et al.*, 2003). Faulting parameters for the Dodecanese Islands, Southern Greece and Southern Ionian Islands events were taken from the Global Centroid Moment Tensor Database (*www.globalcmt.org/CMTsearch.html*), with subsurface rupture lengths estimated using equation 4.8. Both the catalogue of ruptures and of hypocentres can be seen in Figure 4.3.

The Kythera event is the only intermediate-depth event included in the rupture catalogue. The initial focal depth was estimated to be approximately 60 km. Subsequent moment tensor inversions suggest a range of 55 to 65 km, with aftershocks possibly as shallow as 45 km (Konstantinou *et al.*, 2006). The depth cut-off used for the Aegean earthquake catalogue is 60 km, clearly making the Kythira event a borderline case for inclusion. It has been included here for two reasons. Firstly, given the magnitude of this event and the depth range of the aftershocks it is clear that the rupture associated with this event penetrates well into the seismogenic crustal depth assumed previously (60 km). Secondly, although an intermediate depth event, damage from this event was significant and widespread, with an epicentral intensity of VIII on the Modified Mercalli Scale (Konstantinou *et al.*, 2006). Clearly this event is significant for hazard analysis in the Aegean region; the context in which the K-means algorithm is applied.

4.5 Applying K-means Cluster Analysis to Aegean Seismicity

4.5.1 Point Source (Hypocentres)

K-means cluster analysis can be readily applied to the distribution of hypocentres within a region. Here, it will be applied to shallow seismicity; that is seismicity recorded at depths less than 60 km. In the previous chapter it was shown that intermediate and deep earthquakes can be well-modelled by zones corresponding to the location of the subducting slab beneath the south Aegean Sea. For shallow seismicity, K-means is applied in three dimensions of Euclidean space. Hypocentres are referred to Euclidean space by converting each point from a latitude, longitude and depth to a distance in km in the x-direction (with east positive), y-direction (with north positive) and depth in the z-direction. The origin of the Cartesian frame of reference is the centroid of the entire data set.

It is often convenient to think of earthquakes as point sources, and indeed for the purposes of traditional cluster analysis it is necessary. For smaller earthquakes (typically

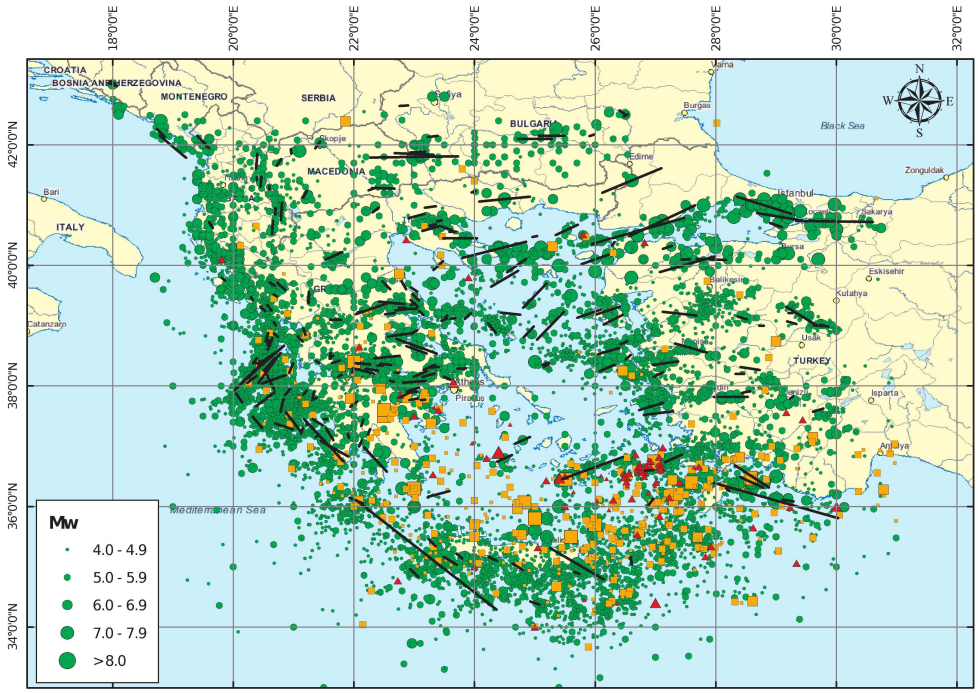

Figure 4.3: Earthquakes in the Aegean region (Burton *et al.*, 2004a). Shallow earthquakes (Depth $<$ 60 km) indicated by green circles, intermediate depth earthquakes (60 $<$ Depth (km) $\leq$ 120 km) by orange square and deep earthquakes (Depth $>$ 120 km) by red triangles. Black lines indicate the location and strike of the ruptures in the Papazachos *et al.* (1999) catalogue, with lengths calculated using equation (Wells and Coppersmith, 1994)

$M_W < 5.5$) the point source approximation is not unreasonable. Larger earthquakes, however, may have rupture lengths extending tens to hundreds of kilometres. To define the sources of these events as points in the same manner as for small events is to ignore one of the most fundamental physical properties that define large earthquakes. Some attempt should be made therefore to take into account the influence of the large events on the K-means process.

The most significant change that can be made to the K-means algorithm without altering the procedure is to use the weighted centroid of the within-cluster data, as opposed to the mean. The weighting of each datum is related to the size of the earthquake. For a cluster with n elements x_i, associated with a weight w_i, the weighted centroid is determined thus:

$$\bar{\mathbf{x}} = \frac{\sum_{i=1}^{n} \mathbf{w}_i \mathbf{x}_i}{\sum_{i=1}^{n} \mathbf{w}_i} \tag{4.9}$$

This is a simple modification that can be readily applied in Euclidean space. It will have the impact of pulling the cluster centroids towards the largest events in the cluster. This may accelerate the convergence of the K-means algorithm toward stability.

The obvious question arises as to what facet of earthquake size should form the basis for the weighting scheme. Earthquake size can be measured in terms of a magnitude unit (here M_W), seismic moment or by dimension of the fault (length, area, slip etc). The choice of weighting metric should not be arbitrary. Those suggested can span a range of scales, which may have an appreciable influence on the behaviour of the clustering algorithm. Here, the preferred weighting metric is fault length. This is chosen as it is consistent with the Euclidean space in which clustering is taking place, and it also scales over a sufficiently great range (1 - 300 km) as to elucidate stronger events from smaller ones. It is not so great a range as to place nearly all the weight in a small number of large events, as would occur with moment weighting. Once again, subsurface rupture length, calculated using the Wells and Coppersmith (1994) relation, is used as the measure of fault length.

The K-means algorithm can only be applied to the complete part of the catalogue. This is simply because the definition of completeness magnitude makes explicit the assumption that not all earthquakes below that magnitude are recorded. As such, the earthquake catalogue contains an incomplete record of events below the completeness magnitude, which may bias the partition. However, completeness magnitude is temporally and spatially variable (Wiemer and Wyss, 2000; Wössner and Wiemer, 2005), depending on the quality of the seismic recording network that has existed in a region. The issue of spatial completeness is not trivial, yet it is hard to distinguish between the variability in completeness owing to spatial differences in recording quality, and genuine spatial variation in the catalogue. Computational methods of assessing completeness do not make the distinction. As such, it is necessary to choose a conservative estimate for completeness magnitude and treat all spatial variability as representative of seismicity.

4.5.2 Line Source (Ruptures)

The weighted centroid in cluster analysis has the advantage of being readily compatible with the K-means procedure, requiring only minor modification. This also has the advantage that it is compatible with existing validity indices used in cluster analysis. It is still treating earthquakes as point sources, albeit ones with varying masses. This is still not representing the true nature of the physical sources of the large earthquakes. What is needed is a modification of the K-means clustering algorithm that is able to integrate the varying sizes of faults, spatial locations and strikes. Such a modification is more complicated. Presented here is a variation of the K-means algorithm, which is designed to partition a set of lines as spatial analogues for fault ruptures. This is the line K-means algorithm described by the flowchart in Figure 4.4.

The data set being input into the line K-means algorithm is the extended rupture catalogue of Papazachos *et al.* (1999) (extended in the current work). Each rupture in this catalogue is given in terms of time, epicentre, focal depth (where constrained, 15 km oth-

erwise), M_W, strike, dip, rake and reference. From this a rupture plane can be defined. Several assumptions have to be made. The first is that the epicentre is found in the middle of the rupture. This is not always true, but with no other information from which to constrain the relative position of epicentre and rupture, then consistency of application is preferable. The second assumption is that the length of the rupture is equal to the 50^{th} percentile value of the Wells and Coppersmith (1994) empirical relation. Again, this approximation has to be made to allow for representation of the ruptures in Euclidean space. This is a source of considerable uncertainty.

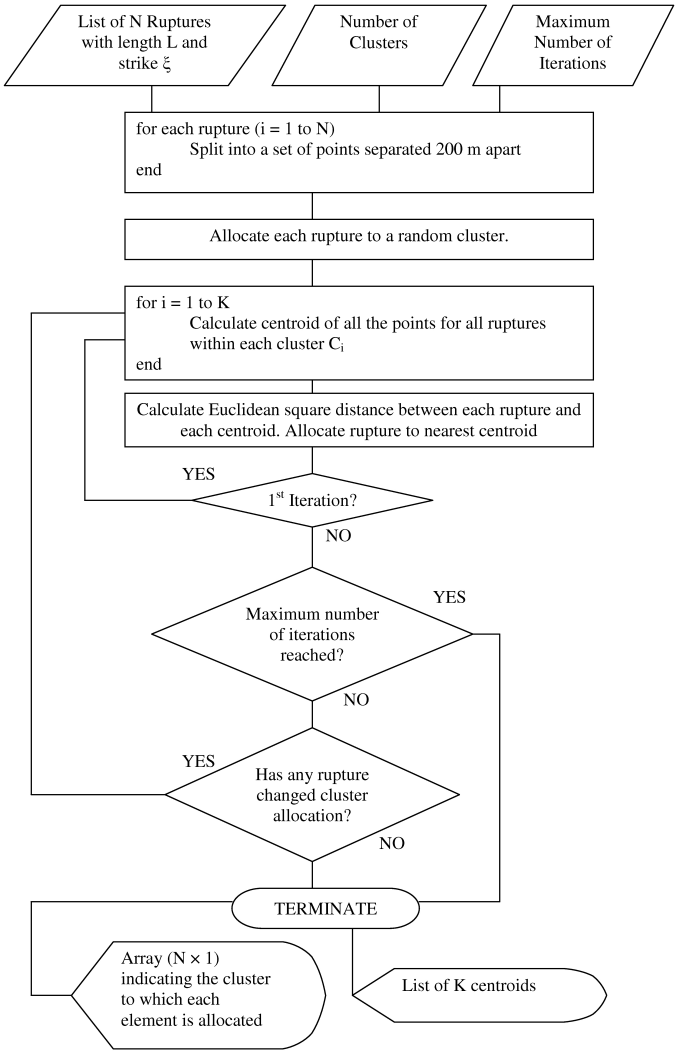

Figure 4.4: Flowchart of the Line K-means algorithm

With the parameters given in the Papazachos *et al.* (1999) catalogue it is only possible to define the linear trace of the rupture. It is not possible to constrain, fully, a rupture plane in three dimensional space using the parameters given. Even where rupture associated with historical earthquakes has been defined it is not possible to elucidate the fault plane from a hypothetical nodal plane with any consistency. The dip direction of many of the normal faulting earthquakes in the Aegean are not known, and both possible planes are consistent with the principal directions of stress. As it is not possible to define a rupture

plane, only a rupture line segment, the line K-means algorithm is only applied in two dimensions here. Whilst it is not possible to consider three dimensional analyses here, in other regions where dip direction can be defined the line K-means algorithm may be extended to consider the partitioning of bounded planes. This would of course come at greater computational expense.

As it is linear segments that are being considered in the line K-means algorithm, several issues arise relating to the compatibility with the traditional K-means process. Considering only the two-dimensional case, two particular issues stand out. The first is how to define the centroid of a set of line segments. The second is how to calculate the distance between two line segments. The latter does not affect the operation of the K-means algorithm, but several cluster quality indices are functions of W(K) and hence require segment to segment distances to be calculated within each cluster.

As is seen in Figure 4.4, the centroid of a set of lines has to be found by discretising each line segment into a finite set of points. Since the ruptures are converted into Cartesian coordinates, much like the point sources, each line is then defined as a set of points separated at a spacing of 0.2 km along its length. Several spacings were considered and 0.2 km offered the optimum compromise between preserving the length of the faults and over-scaling the largest events. It is the centroid of all the within-cluster points that is determined.

The distance between two line segments is defined as the shortest Euclidean square distance between the two segments. This is illustrated in Figure 4.5. Should line segments intersect, the distance between them is equal to zero. The method for calculating Euclidean square distance between two line segments is that of Allen *et al.* (1993).

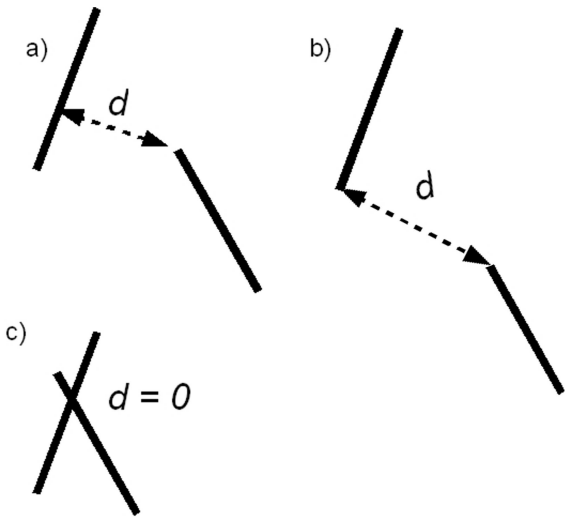

Figure 4.5: Illustration of the line to line distance definition

Only 223 ruptures are used to define the spatial distribution of faults across the entire Aegean region. Here these are effectively only pseudo-ruptures, since they are subject to

uncertainty in length and location. Owing to the magnitude-length scaling of the fault ruptures, it is the largest earthquakes that dominate the delineation of faults, and consequently the cluster analysis. In some parts of the Aegean, smaller ruptures are distributed around larger ruptures in a compact manner, possibly indicating repeated events on the same rupture or activity on branch faults connected to the larger rupture. Elsewhere, smaller ruptures are well distributed, possibly indicating uniform zones of small to moderate seismicity.

Much like the point source application of K-means, the issue of completeness in the rupture catalogue is not a trivial one. From a catalogue of only 223 events, computation of M_C is misleading. The catalogue of ruptures is not complete in the same sense as that of the earthquake catalogue; comparison is futile. It is also recognised that the largest events will dominate the spatial distribution of ruptures. This can mean that the impact of lowering or raising the threshold magnitude of the rupture catalogue may not have a significant impact on the partition. Since the ruptures are intended to represent more permanent features of seismotectonics in the Aegean region, then it cannot be inferred that they are temporally variable in the manner that seismicity is. Instead, the ruptures are a representation of a database of information as yet incomplete. Consequently all ruptures in the catalogue are included in the analysis, though it is recognised that line K-means algorithm should be repeated as more ruptures are recorded.

4.5.3 The Optimum Set of Seeds

The optimum partition for a given value of K is found from an ensemble of 100 partitions. The partition producing the lowest TWCSS within the ensemble is selected as the optimum partition. For each member of the ensemble the starting cluster centres are determined using RP initialisation, with n/K data points allocated to each cluster (the remainder allocated at random).

4.5.4 Optimum K

This is perhaps the most difficult parameter to determine. To assist in the identification of a partition or partitions most appropriate for the spatial distribution of seismicity in the Aegean region, several indices were compared. Several indices were tested on synthetic sets of clustered data in three dimensions. The synthetic data sets contained a predefined number of compact clusters. Those indices that correctly identified the number of clusters in the data set were then used for the Aegean catalogue.

It should be noted, that most of the indices selected do not test the null (K = 1) case. The implication of the null case, from a seismotectonic perspective, is that seismicity is uniformly distributed across the entire Aegean region. By considering seismogenic source delineation to begin with, this notion is already rejected. This is done not on the

basis of the cluster analysis, but on the basis of existing knowledge of seismicity and tectonics in the Aegean.

The indices that identified the correct number of clusters in the synthetic data set were:

1. Xie and Beni (1991) Index (XB)

$$XB(K) = \frac{\sum_{i=1}^{K}\sum_{j=1}^{N}(\mu_{ij})^2 (x_j - m_i)^2}{N \min_{i,j}\{x_j - m_i\}^2} \tag{4.10}$$

   Where m_i is the centroid of cluster C_i and μ_{ij} is the element of a N $\times$ K matrix that takes the value of 1 if x_j is within cluster C_i and 0 otherwise. The optimum K is that which minimises XB. It should be noted that the term in the denominator refers to the minimum distance between a datum and its allocated centroid for the entire data set. Consequently, if a singleton cluster is produced XB is infinite. To trap this error XB assumes a value of 1×10^{15}, which is well outside the feasible range of values. This index does therefore assume that a partition producing a singleton cluster is not an optimum, and therefore that the specific value of K does not define the number of clusters in the data set appropriately.

2. Calinski and Harabasz (1974) Index (CH)

$$CH(K) = \frac{B(K)/(K-1)}{TWCSS(K)/(N-K)} \tag{4.11}$$

   The optimum K is that which maximises the Calinski and Harabasz (1974) index. Since the between cluster sum of squares is a numerator in this index, it clearly improves with cluster separability.

3. Silhouette index (Kaufman and Rousseeuw, 1990) - equation 4.7

   The optimum number of clusters is that which minimises the mean of the silhouette index. This index is not implemented in the line K-means algorithm owing to computational expense.

4. Krzanowski and Lai (1988) Index (KL)

$$KL(K) = \left| \frac{DIFF(K)}{DIFF(K-1)} \right| \tag{4.12}$$

$$DIFF(K) = (K-1)^{2/d} WK_{K-1} - K^{2/d} WK_K \tag{4.13}$$

   d refers to the dimension of the data set. The optimum K is that which maximises KL.

Also tested was the "gap statistic" (Tibshirani *et al.*, 2001). This validity index has the advantage of being able to test the null case. It proved successful at identifying clusters

that were very will separated, but the performance dropped drastically as the separability of the clusters worsened. It also has the disadvantage that it applies K-means to large sets of randomly distributed synthetic data, as well as the observed data being clustered. This results in a large increase in computation, which is not well compensated by performance in identifying the correct number of clusters.

All of the indices described previously may vary in performance depending on the nature of the data set being used (Krzanowski and Lai, 1988; Tibshirani *et al.*, 2001). In particular, where clusters are not well separated it may be that these indices are sensitive to small changes in the partitions. This is where the possibility of error emerges from not necessarily having the global optimum partition. Experience indicates that whilst many of these indices are consistent when the experiment is repeated, others can be subject to change if a better partition for a particular K is found on the repeat.

Although it is recognised the optimum K is that which maximises or minimises a particular index, it is likely that several maxima or minima may emerge. This can be addressed in many ways. Firstly it may be necessary to increase the number of K-means runs in the ensemble analysis. This may produce a better partition for some of the indices. Those maxima/minima remaining robust to an increased number of trials are likely to be representative of the optimum K. If several optima are apparent then it may become necessary to examine the partitions produced. It may be the case that more than one value of K could represent an appropriate partition. If this is the case then it may be worth producing several different source models in the seismic hazard analysis, and weighting them in a logic tree via the strength of the partition index. Should several localised optima emerge, each of the corresponding partitions will be considered.

One last issue to consider in the assessment of optimum K is the maximum K to be tested. In some applications of cluster analysis it is not uncommon for K to be tested in the range $2 \leq K \leq N$. Alternatively, it is suggested by Sheng and Liu (2006) that an appropriate upper limit for K should be $\approx \sqrt{N}$. In practise the choice of maximum K is often arbitrary, normally a conservative value reflecting the likelihood of a particular upper limit to the number of useful clusters in a data set. The choice of maximum K can reflect the needs of the user. If one is interested in partitioning data in an unsupervised capacity with no further requirements necessarily made of the clusters, then $K_{MAX} \rightarrow N$ may not be inappropriate. This does come at a cost. There is an increased likelihood of singleton clusters, which will bias some of the validity indices.

In this application partitions of the earthquake catalogue are used as a means of analysing spatial patterns of seismicity, ultimately with the intention of delineating seismic source zones. The properties of individual source zones need to be considered, and some discussion has been given in Chapter 3. In particular, there arises the issue as to how reliable the parameters of earthquake recurrence are when the zone contains only a very small number of earthquakes. Whilst clusters containing only a small number of events,

even singletons, are tolerable in unsupervised cluster analysis, they serve little purpose as seismic zones. They will contain too few events from which to constrain earthquake recurrence parameters. Essentially they risk over-fitting the data and inferring a greater degree of spatial variability than may actually exist. Also, clusters with very few events are more likely to represent transient features of seismicity in the catalogue, thus risking invalidation of the assumption of Poissonian seismicity.

In this analysis the earthquake catalogues tested contain approximately 1700 to 2500 events, depending on the subset being considered. Following the example of Sheng and Liu (2006) where $K_{MAX} \approx \sqrt{N}$, K is tested in the range $2 \leq K \leq 50$. For the line K-means algorithm singleton clusters are more tolerable since many earthquakes can be related to one seismogenic source. Since $K_{MAX} \approx \sqrt{N}$ would produce a K_{MAX} of approximately 15, then it is clearly worth expanding the range beyond this limit. Ocular analysis of the spatial distribution of ruptures would suggest that there are many locations where ruptures cluster spatially. Consequently $K_{MAX} \rightarrow N$ extends the range beyond that which is necessary. The decision is taken to implement the line K-means algorithm for the range $2 \leq K \leq 35$. This represents a compromise between the tolerance of clusters with few elements and the limits of the small number of data in the catalogue.

4.5.5 Performance of the Validity Indices - Earthquake Hypocentres

The K-means algorithm is applied to six subsets of earthquakes from the combined Aegean catalogues of Burton *et al.* (2004a) (20th Century), including the supplementary period 2000 to 2005 (homogenised from the NOA catalogue), and Papazachos and Papazachou (1997) (Pre-20th Century).

1. Subset 1: 20th Century Events - Shallow (Depth $\leq$ 60 km), $M_W \geq 5.2$.
2. Subset 2: As subset 1, declustered using Reasenberg (1985).
3. Subset 3: Pre-2000 AD - Shallow, $M_W \geq 5.2$.
4. Subset 4: As subset 3, declustered using Reasenberg (1985).
5. Subset 5: 1900 AD - 2005 AD - Shallow, $M_W \geq 5.2$.
6. Subset 6: As subset 5, declustered using Reasenberg (1985)

For each of these subsets, the variation in validity index with K is plotted (Figures 4.6 to 4.11). Four indices are shown: the inverse of the Xie and Beni (1991) index (a), Silhouette statistic (b), Calinski and Harabasz (1974) index (c) and the Krzanowski and Lai (1988) index (d). The inverse of the Xie & Beni index is displayed as that is the only index for which the optimum K is identified as a minimum. By using the inverse index those values

of K that produce a better fit are elucidated. Therefore, in each figure it is the maxima of each index that are identified as optima of the number of clusters.

It is readily apparent from the indices shown in Figures 4.6 to 4.11 that the optimum K identified differs considerably depending on the index used. Despite being recognised as a common index for identifying optimum K on some data sets, here there appears to be a strong correlation between the silhouette statistic and K, with the silhouette statistic decreasing exponentially with increasing K. This result is a clear manifestation of the weakness of the silhouette index when clusters are not well separated. This comes from the numerator in the index, where the difference between mean within-cluster distance and mean between-cluster distance diminishes.

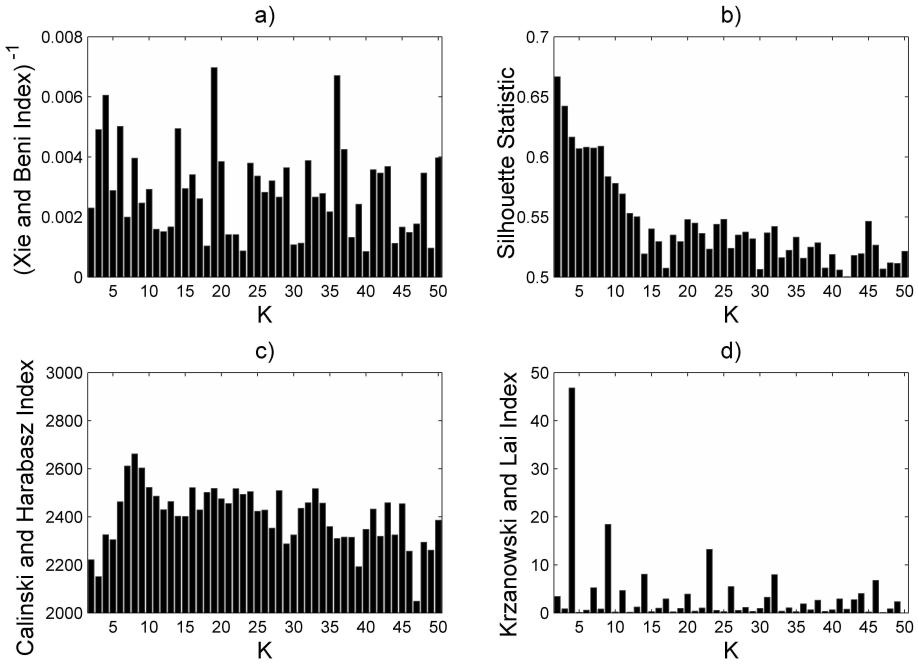

Figure 4.6: Cluster validity indices for partitions of the full 20th century catalogue (Subset 1)

The correlation between silhouette index and K when clusters are poorly separated would suggest that the silhouette index cannot be considered any further as a means of identifying optimum K for these data sets. Closer scrutiny of the Calinski and Harabasz index produces a more complicated dilemma. For almost all the subsets of the catalogue there appears to be rapid increase in the index for low values of K, reaching a maximum consistently in the range $7 \leq K \leq 10$. Beyond this range, Calinski and Harabasz index appears to decay slowly with increasing K, with considerable noise. This leads to one of two scenarios. The first is that the Calinski and Harabasz index is robust to the choice of subset and that the optimum K lays in the range 7 to 10 clusters. The alternative is that there exists a correlation with K that will produce a characteristic peak in this range of K. Although there are some localised maxima for higher K apparent in some subsets, the general shape of the bar plots is quite consistent. Closer scrutiny of the index

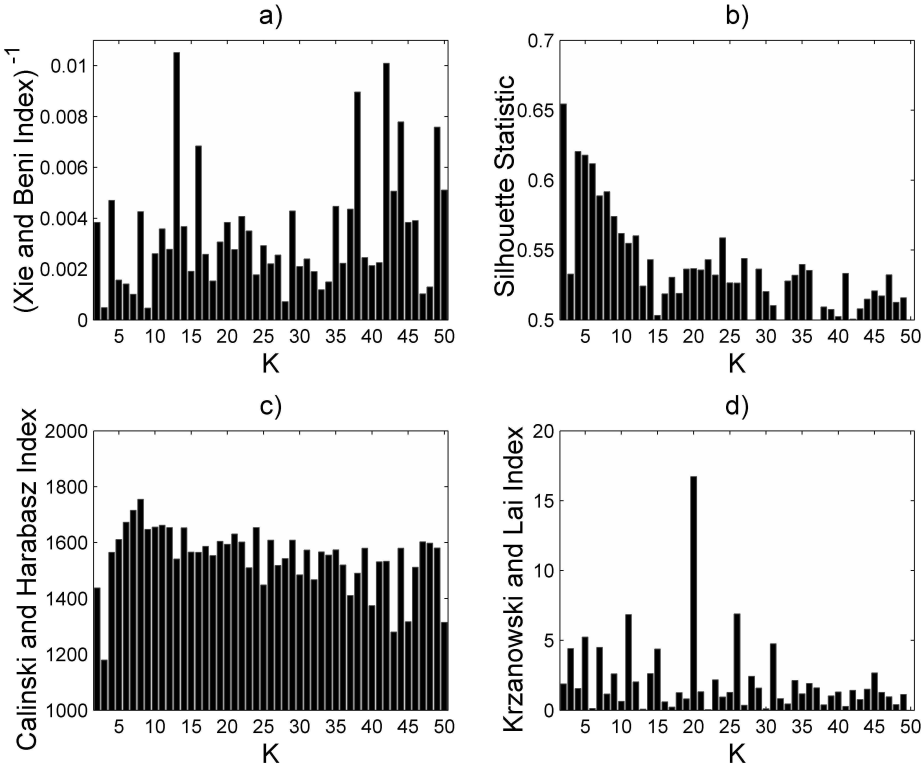

Figure 4.7: Cluster validity indices for partitions of the declustered 20th century catalogue (Subset 2)

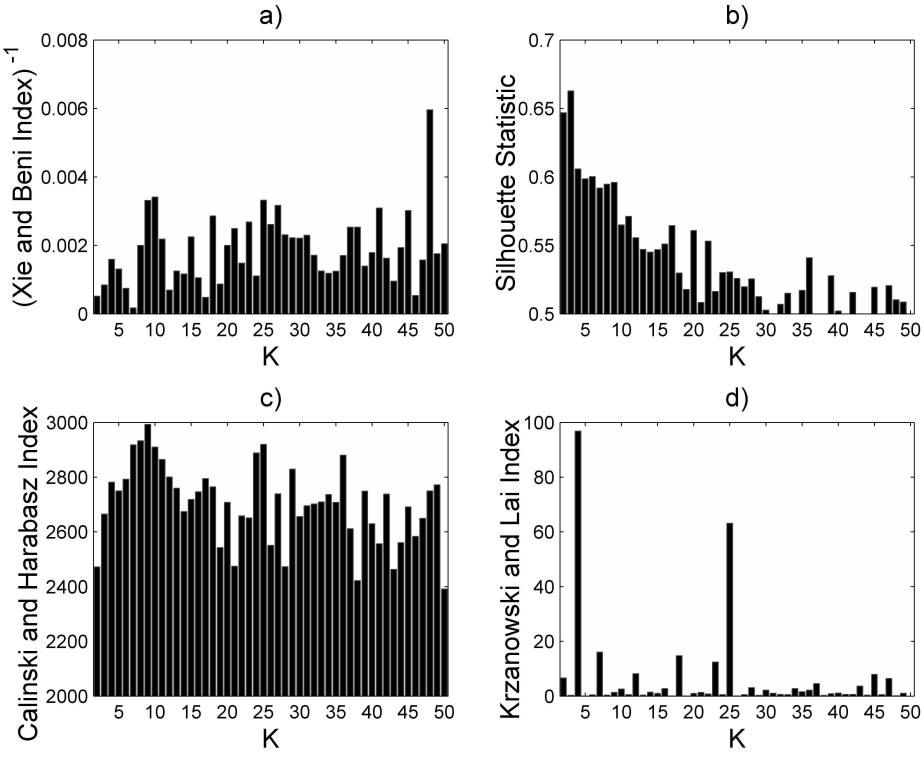

Figure 4.8: Cluster validity indices for partitions of the full catalogue (Pre-2000 A.D.) (Subset 3)

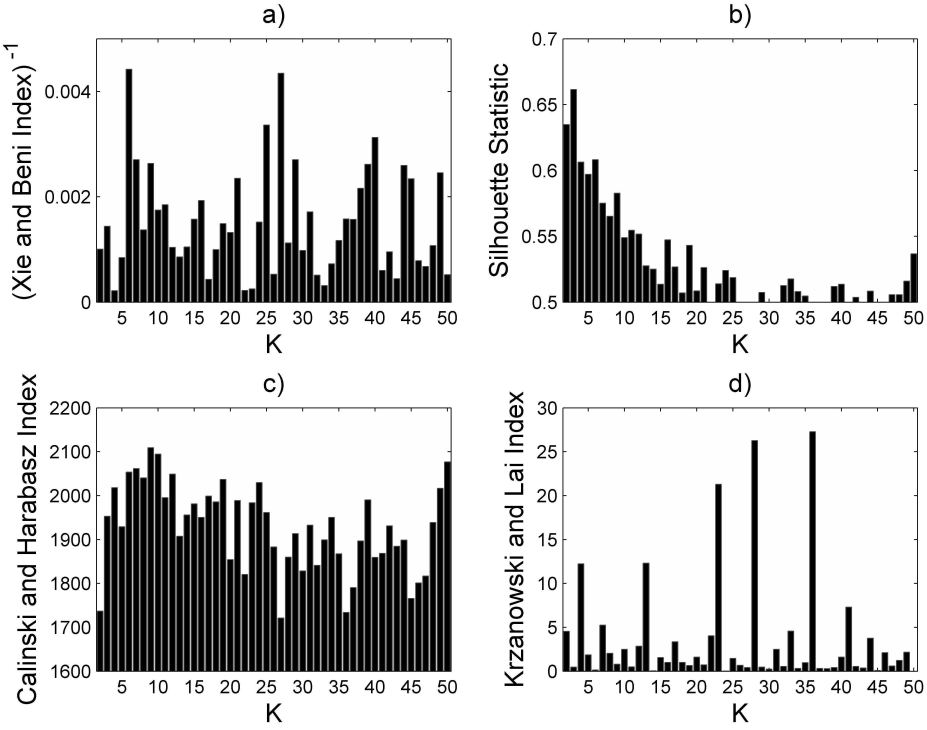

Figure 4.9: Cluster validity indices for partitions of the declustered Catalogue (Pre-2000 A.D.) (Subset 4)

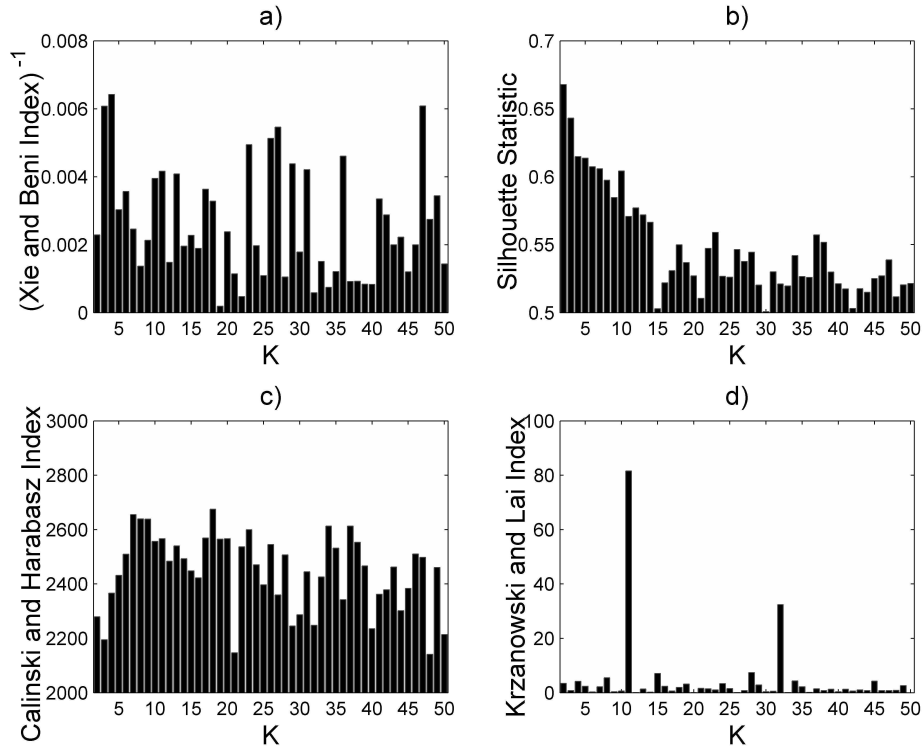

Figure 4.10: Cluster validity indices for partitions of the full catalogue (1900 2005 A.D.) (Subset 5)

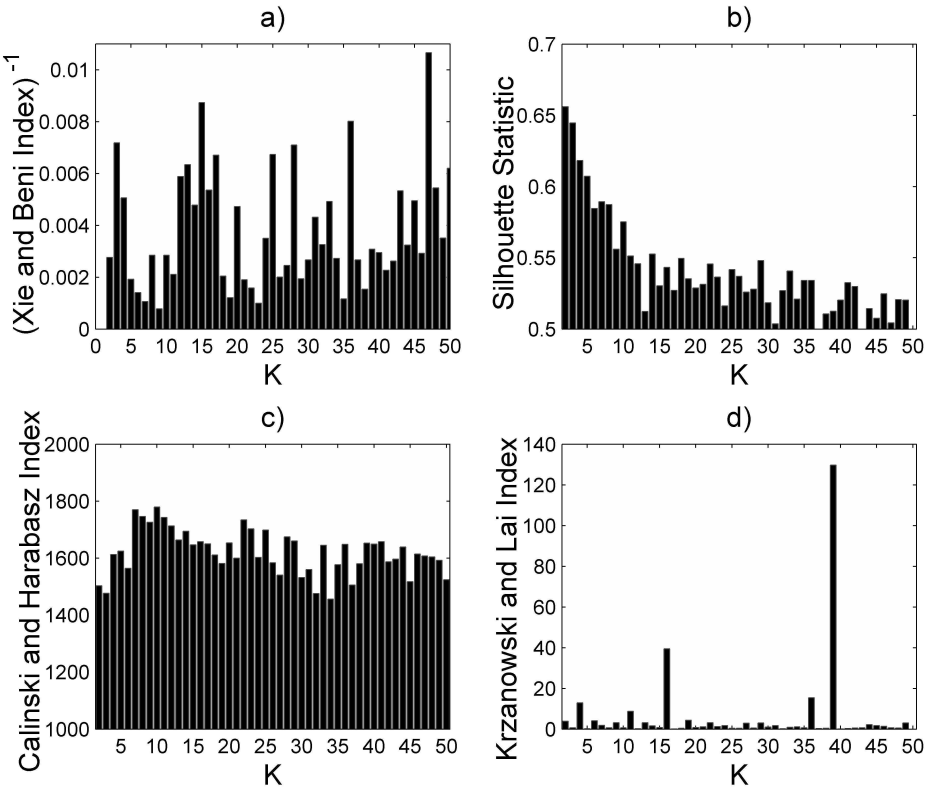

Figure 4.11: Cluster validity indices for partitions of the declustered Catalogue (1900 2005 A.D.) (Subset 6)

itself would suggest that a characteristic shape of the index with increasing K is not unexpected. Given that the index is also a function of between-cluster sum of squares, the separability of clusters influences the identification of optimum K by this index. Nevertheless, since the possibility exists that Calinski and Harabasz index is in fact identifying the optimum K, then the values in the suggested range may have to be given further consideration by scrutinizing the partition in more detail. This is done in section 4.7.

Both the Xie & Beni and Krzanowski & Lai indices share some common features. Firstly, there is no obvious correlation between increasing K and each index. In both indices localised maxima are apparent, but these maxima are sensitive to the subset used. There are very few values of K for which the Xie & Beni index and the Krzanowski & Lai index agree are possible optima. Several values do appear to stand out with some consistency. The first is K = 4, which emerges as a local maximum in the Xie & Beni index for subsets 1 and 5, and in the Krzanowski & Lai index for subsets 1, 3, 4 and 6. Also of interest is K = 36, which stands out in the Xie & Beni index for subsets 1, 5 and 6, and in the Krzanowski & Lai index for subsets 4 and 6. Several values in the range $23 \leq K \leq 26$ also appear with some consistency, with K = 25 standing out particularly strongly in subset 3 (including in the Calinski and Harabasz index). Of the two indices that don't appear to correlate with K, it is the Krzanowski & Lai index that tends to elucidate better fitting partitions from poorer ones.

Whilst both the Xie & Beni index and the Krzanowski & Lai index are able to identify better fitting partitions, the inconsistency is something that needs to be addressed. In both indices the "optimum" partitions are sensitive to the subset of the data used. Is this result detrimental to the intended purpose of the cluster analysis: the delineation of spatially uniform areas of seismicity? One might imagine that in this context the optimum partition is one that would remain robust regardless of which subset of the catalogue is used. Since time-independent seismic hazard assumes stationarity in the seismicity then an appropriate zone model would be one that reflects true spatial variation in seismicity, not just apparent variation within the catalogues. In reality, where clusters are not well-separated, as in these data sets, differences of several hundred events will influence the partition. In particular, it will influence the optimum partition inferred from cluster quality indices.

When comparing the subsets of the catalogue indicated here, an important consideration is the impact of Poisson declustering on the partition, and on the optimum number of clusters identified. It cannot necessarily be presumed that the impact of removing non-Poissonian events will lead to an increase or decrease in the optimum number of clusters identified by the K-means algorithm. For there to be little difference, it would indicate that non-Poissonian events are distributed evenly across the entire Aegean region. The difference in activity between plate margins and intra-plate regions would imply that non-Poissonian events are more likely to be found along the Adriatic-Hellenic plate boundary and the North Anatolian Fault. If non-Poissonian events are not distributed evenly across the Aegean then it is to be expected that removal of such events from the catalogue would result in remarkably different partitions. This is largely borne out by the inconsistency in the optimum K identified when considering the non-Poissonian subsets (1, 3, 5) and the declustered subsets (2, 4, 6). Since it is the Poissonian catalogue that will be used for seismic hazard analysis later on in this work, however, emphasis will be placed on the partitions of the declustered catalogue.

4.5.6 Performance of the Validity Indices - Ruptures

Only one data set is used for the line K-means algorithm. Three indices are used to test the optimum K: Xie & Beni, Calinski and Harabasz, and Krzanowski & Lai. The results can be seen in Figure 4.12.

The first notable observation is that the Calinski and Harabasz index displays a similar trend in correlation with K as it did for the point data. The peak of the curve in this index occurs around K = 10, with a slow decay in the index with increasing K above this value. The main difference in this index is that the actual peak occurs at K = 2. This is a result that has emerged in several indices and one that needs to be treated with some caution. This can be interpreted as an artefact of the cluster validity indices in a data set with poorly separated clusters. If clusters are poorly separated then the validity index

may only be optimised at K = 1. The K = 2 case would then emerge as a near optimum partition, which could very well produce a "better" partition than for any higher K. From an unsupervised learning perspective the analysis would suggest that in the K = 1 or K = 2 case, spatial distribution of epicentres is approximately uniform. If attempting to delineate zones then this result would, taken at face value, imply that seismicity is uniform across the study region. In some regions this scenario could be feasible. In the Aegean our knowledge of the seismotectonics is such that it is easy to determine that seismicity is not uniform across the Aegean. This is one such circumstance where the user judgment is implemented instead of simply accepting the results of any algorithmic approach. Closer scrutiny of the "optimum partitions" is needed before assigning any value of K to be the optimum for hazard analysis.

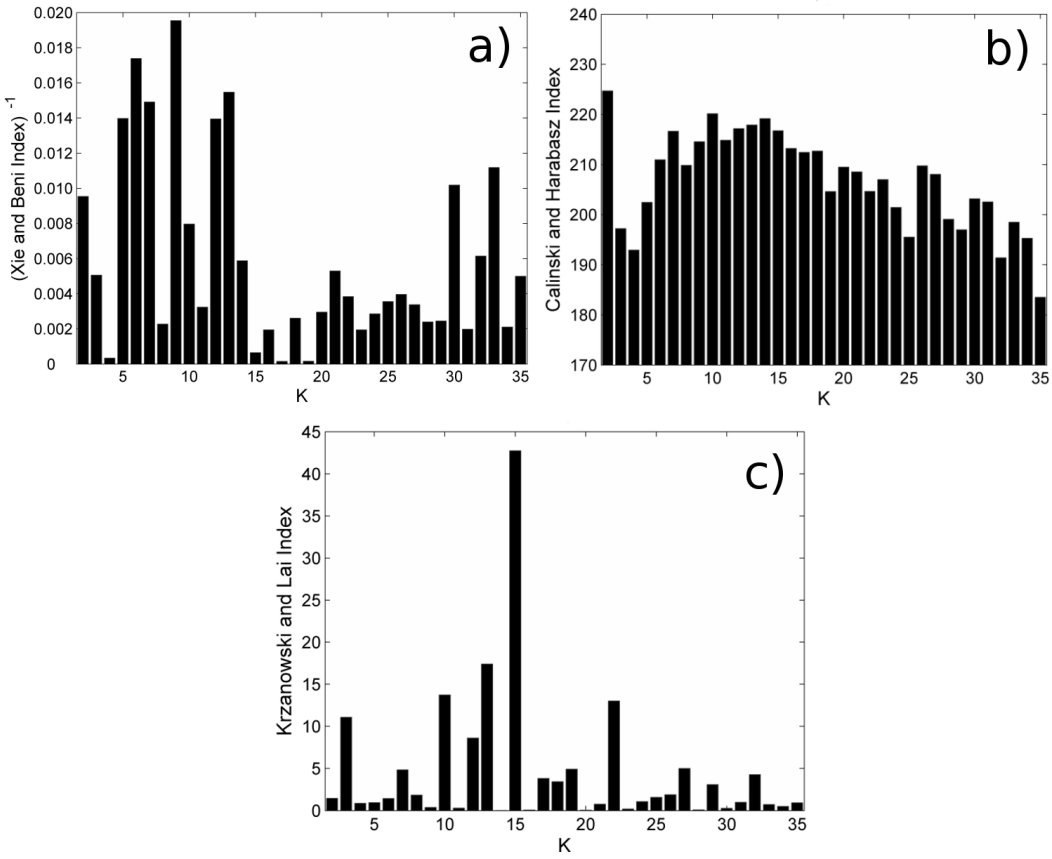

Figure 4.12: Cluster validity indices for partitions of the rupture catalogue

There is more agreement in all three indices for the line K-means partition than for the point K-means partition. In all three indices the maxima appear to be found in the range $10 \leq K \leq 15$. In particular K = 10 is identified as a "good" partition in the Calinski and Harabasz index and the Krzanowski & Lai index, as does K = 15. K = 13 performs well across all three indices. Also, there appear to be a few well-performing values of K towards the higher end of the range, with K = 30 standing out in the Xie & Beni and Calinski and Harabasz index. The greater degree of consistency between these indices for the line K-means partitions suggests a greater degree of confidence in the optimum K values identified. This may partly arise as an artefact of the smaller number of data in

the catalogue itself (223 ruptures as opposed to more than 2000 hypocentres).

4.5.7 Limitations of Validity Indices

The validity indices shown here have identified several values of K that represent "better" partitions of the data input. They have not identified a particular optimum K that remains robust to the input data, as would be an expectation in typical cluster analyses. This result is not surprising, nor does it represent a failure of the K-means approach to identify a partition suitable for delineation into a seismogenic source model. Many indices used to assess partition quality quantitatively are developed under the assumption that the clusters are well-separated. As is shown here, where they are not well-separated many indices may fail altogether, whilst others will identify several possibilities.

A more fundamental issue exists: how to translate the partitions into sources for seismic hazard analysis? An important point that has been raised several times in this discussion is the distinction between a cluster in a data set and a source zone. To define the former, only the spatial distribution need be considered. To define the latter, more consideration needs to be given to the properties of seismicity within the cluster. Of particular importance is the number of earthquakes, the magnitude range, the maximum magnitude and parameters describing earthquake recurrence. The Cornell (1968) - McGuire (1976) method of PSHA requires that for each seismogenic source (here a zone) a seismicity rate, b-value and M_{MAX} are defined. Singleton clusters or clusters with very few events may be tolerable when simply considering the partition of the data. When used to delineate source zones they are simply unsuitable as they contain too few data from which to define other seismicity parameters. Consequently, where the K-means algorithm and respective validity indices may define an optimum partition, this does not immediately translate into the optimum source model. What is needed therefore is a means of identifying optimum K, related to the context of seismic hazard.

4.6 Assessing the optimum partition using stochastic seismic hazard analysis

4.6.1 The Use of Stochastic Seismic Hazard Analysis

To identify, both quantitatively and objectively, the seismicity-dependent zone models that are more suitable for the region under consideration, it is appropriate to assess how they perform in a seismic hazard analysis. The Monte Carlo method of PSHA (Shapira, 1983; Musson, 1999b; Giardini *et al.*, 2004) is a useful technique for testing the source dependence in seismic hazard analysis. A basic procedure is suggested in Musson (2004) whereby synthetic earthquake catalogues are produced using the zone model and forecasts are compared to those from the observed earthquake catalogue by χ^2 analysis. A

similar approach is adopted here. It is extended further in that the χ^2 analysis is derived from the expected ground motion over a set of data points.

For a set of partitioned earthquakes, synthetic earthquake catalogues can be created in different ways. On the assumption that the earthquakes used in the K-means analysis are Poissonian, the parameters of each synthetic earthquake can be selected in the following ways. Hypocentres can be randomly sampled (with replacement) from the observed dataset (Ebel and Kafka, 1999), or from a uniform source zone (Musson, 1999b; Giardini *et al.*, 2004). Magnitudes can be sampled from the observed dataset or from the cumulative distribution function (CDF) of the Gutenberg and Richter (1944) relation bounded at upper (M_{MAX}) and lower magnitudes (M_{MIN}) (Kramer, 1996):

$$F\left(M_w\right) = \frac{1 - \exp\left[-\beta\left(M_W - M_{\min}\right)\right]}{1 - \exp\left[-\beta\left(M_{\max} - M_{\min}\right)\right]} \tag{4.14}$$

where $\beta = b \ln\left(10\right)$.

When synthetic catalogues are created by random re-sampling of the observed catalogue (Ebel and Kafka, 1999) then a partitioned dataset is useful. For example, Ebel and Kafka (1999) created synthetic catalogues for the North-eastern United States by randomly re-sampling from the entire catalogue covering the whole study region. This assumed that seismicity is effectively homogeneous, at least in terms of fault mechanism and strain rate. When considering a region as tectonically diverse as the Aegean, this assumption is less reasonable. The K-means partition allows for the regional catalogue to be broken down into smaller subsets. Creating synthetic catalogues by re-sampling earthquakes within each subset ensures that the synthetic catalogue maintains the same regional differences in earthquakes as the observed dataset.

A set of points partitioned as one cluster is not akin to an areal or spatial zone. Using the K-means partition to develop a model of uniform seismic source zones is a more complex matter. In this analysis, and for the purposes of automation, uniform zones are created by partitioning (without weighting) the entire source region around the centroids of each cluster, in a manner similar to Voronoi tessellation around the K-means centroids, within a finite region (Du *et al.*, 1999). This is achieved by creating a grid of discrete points (spaced at $0.02^\circ \times 0.02^\circ$) across the Aegean region. Each grid point is then allocated to its nearest cluster centre, using Euclidean square distance as the metric. The zone is then represented as the collection of grid points allocated to the respective centres. In the Monte Carlo analysis, synthetic epicentres within a zone are generated by random sampling, with replacement, the grid points within each zone. To create a zone from the grid points each sampled epicentre is then adjusted by adding random scatter around the sampled grid point.

The tessellation method of zone delineation, as implemented, here is arguably the most simplistic approach to the problem of translating partitions into source zones. It has the

advantage of transparency and computational efficiency. The former is particularly important if comparing several models in an analysis of epistemic uncertainty. The grid points are allocated to each centre with a uniform weighting. This does not necessarily have to be the case as the centroids could be weighted by number of events or within-cluster average moment or magnitude. Tessellation is, however, a simple approach, which will delineate zones in a mosaic pattern within the region being considered. This may appear counter intuitive as it may often be that case that source zones do not appear elongated along the largest ruptures in the zones.

Regions where seismicity is trivial (i.e. no earthquakes greater then M_C are recorded in the catalogue) are excluded from the source model to avoid extending zones well offshore into regions of very low seismicity (e.g. the Black Sea, western Libyan Sea). This automated zonation process is used for comparison of source models with the number of zones in the range $2 \leq K \leq 50$.

Alternatively the creation of geometric zones of uniform seismicity around these clusters may be done by inspection, using the partitions as a guide. This allows the analyst to create zones that may reflect more accurately the seismotectonic variability of the region, whilst preserving the partitions of seismicity identified by the K-means cluster analysis.

To assess, quantitatively, how well the source model represents the observed seismicity of the Aegean, the region is overlaid by a new set of NG grid of points. Using the earthquake catalogue of Burton *et al.* (2004a) (cut off at the magnitude of completeness) the maximum ground motion observed in the period 1900 to 1999 is calculated for each grid point from an appropriate attenuation relation. The source model is then used to simulate 100 years of seismicity and the maximum ground motion at each grid point determined for the 100 years of synthetic data. This simulation and calculation is repeated a large number (Nsyn = 100) of times and the spread of maximum observed ground motions for the different 100 year simulations is developed. Using the geometric mean ($\overline{X}$) and standard deviation (σ) of the N_{SYN} maximum simulated ground motions in 100 years at each grid point, it is possible to calculate a χ^2 for the model:

$$\chi^2 = \frac{1}{NG} \sum_{i=1}^{NG} \frac{\left(\max\left(Obs\left(\log_{10}\left[SGM_i\right]\right)\right) - \bar{X}_i\right)^2}{\sigma_i^2} \tag{4.15}$$

To allow for spatial variation in the performance of the model, the normalised difference between the observed and expected ground motion is used. Models producing a lower χ^2 value indicate a better fit to the observed data.

It should be acknowledged that variability in χ^2 originates from two sources. This first is the true variability with K, which is the focus of the analysis. The second source comes from the inherent variability within the Monte Carlo technique. The σ term within the χ^2 definition adjusts for the aleatory uncertainty that is captured in the Monte Carlo method.

In spite of this, some of this uncertainty still propagates into the χ^2 value itself. Where the difference in fit between two values of K is small it cannot be said that the fit of the two values is significantly different.

4.6.2 Selection of the Preferred Ground Motion Parameter

Traditional methods of seismic hazard analysis use Peak Ground Acceleration (PGA), Peak Ground Velocity (PGV) and Peak Ground Displacement (PGD), and their spectral ordinates, as the preferred strong motion parameters. For the χ^2 calculation here, ground motion quantified by a single parameter (e.g. PGA) is needed. Whilst PGA may seem the more likely candidate, for the purposes of assessing maximum ground motion for a 100 year period, it has some shortcomings. In stochastic seismic hazard analysis, the greatest PGA can come from small earthquakes in the near-field, even if such earthquakes may not be especially damaging. Instead, a duration-dependent parameter of strong ground motion is used: Arias Intensity (I_a) Arias (1970) defined as:

$$I_a = I_{xx} + I_{yy} = \frac{\pi}{2g}\int_0^{t_0} \left[a_x\left(t\right)\right]^2 dt + \frac{\pi}{2g}\int_0^{t_0} \left[a_y\left(t\right)\right]^2 dt \tag{4.16}$$

where a_x and a_y are the horizontal accelerations in the x and y directions respectively, t_0 is the duration of strong shaking and g is the acceleration due to gravity ($\approx 981\ cm\ s^{-2}$).

The attenuation relation used is that of Danciu and Tselentis (2007), which is derived from 335 strong motion records from 151 Greek earthquakes. The strong motion records span an epicentral distance of 1 - 150 km, and a moment-magnitude range of $4.5 \leq M_W \leq 7.0$.

$$\log_{10}\left(I_a\right) = -2.663 + 1.125M_W - 2.332\log_{10}\sqrt{R^2 + 13.092^2} + 0.028S_o + 0.200F_o + 0.524\sigma \tag{4.17}$$

where, R is epicentral distance, S_0 indicates engineering soil class (0 for rock, 1 for stiff soil, 2 for soft soil), F_0 indicates fault mechanism (0 for normal faulting, 1 for strike-slip/thrust faulting), and σ the total standard error including inter- and intra-event variability. An alternative attenuation relation for Arias Intensity is that of Travasarou *et al.* (2003), which is developed from a non-linear regression on a larger global dataset of earthquakes (including some from Greece and Turkey). A crucial difference between this and the Danciu and Tselentis (2007) attenuation relation is in its treatment of non-linear variation in $\log\ I_a$ with magnitude. For the purposes of seismic hazard calculation the Travasarou *et al.* (2003) relation may be preferable to reduce the impact of near-field small events. Here the focus is on comparing the performance of source models. As such, a linear scaling of $\log\ I_a$ with magnitude is acceptable, hence the Danciu and Tselentis (2007) relation, with its emphasis on Greek earthquakes, is chosen.

4.6.3 Application to point data

In the stochastic seismic hazard analysis presented here, the partitions are the same as those used Section 4.5.5. Only the 20th century catalogue (subsets 1 and 2) is used for the convenience of simulating 100 years of seismicity. Minimum magnitude used in the hazard analysis therefore is M_W 5.2, the completeness magnitude for the period 1900 to 1999. Although it is the Poisson declustered catalogue that conforms to the standard of seismic hazard analysis, the χ^2 analysis for the full catalogue is presented too. Comparison between the two types of catalogue may help in the identification of more stable partitions. The zoned and non-zoned seismic hazard analyses are shown in Figure 4.13.

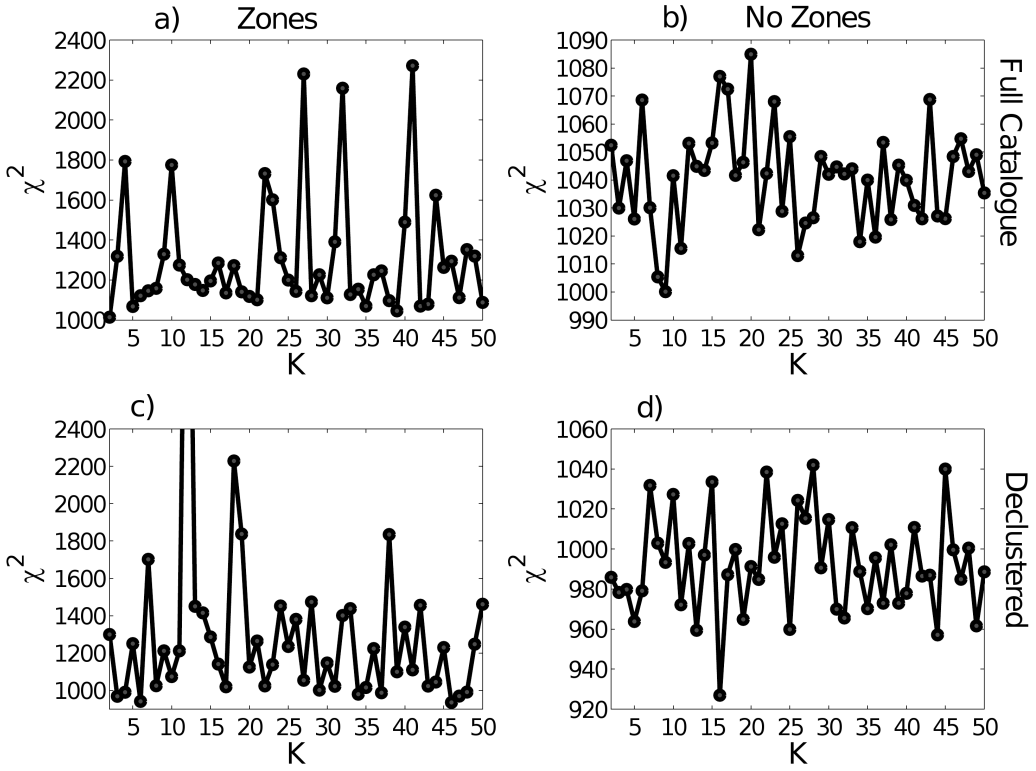

Figure 4.13: Variation in χ^2 value with number of clusters using random re-sampling (b and d) and uniform zones (a and c). The catalogue used is the Aegean 20th century catalogue with only shallow (Depth $<$ 60 km) events complete at M_W 5.2. a) and b) use the full catalogue (Subset 1), c) and d) the declustered catalogue (Subset 2)

It is immediately apparent that there is considerable variation in χ^2 when uniform zones are used. Generally, uniform zones produce a higher χ^2. This is because the seismicity is smoothed out across the zone. When zones are not used, the synthetic epicentres are simulated by random re-sampling of the observed epicentres, hence simulated seismicity will resemble the observed seismicity more strongly and χ^2 will be lower. There are some partitions where the zoned model produces a lower χ^2 than the non-zoned approach. For the Poisson declustered catalogue this occurs at K = 3, 6, 22, 34, 46 and 47, and for the full catalogue this occurs only at K = 2. Given the inherent variability within the Monte Carlo process, the differences in χ^2 between the zoned and non-zoned approach are not

statistically significant. They may suggest that these particular partitions are possibly more appropriate for the delineation of seismic source zones. That said, the observation that there are more partitions for which zoned and non-zoned approaches give similar results when using the declustered catalogue, may suggest that the quality of fit is an artefact arising from the fact that there are fewer earthquakes in total in the declustered catalogue than in the full catalogue.

4.6.4 Application to rupture data

To undertake a seismic hazard analysis comparison when using partitions of rupture catalogues, it is obviously still necessary to use an earthquake catalogue for calculation of recurrence parameters. To allow for comparison with the point source approach, only the 20th century catalogue, both full and Poisson declustered (subsets 1 and 2), is used. Observed earthquakes are partitioned into clusters around the centroids of the rupture partitions. Uniform zones are delineated in the same manner as for the point source data. Here it is the rupture centroid that is used as the focus of the Voronoi delineation, and not the point centroid. It is possible in this automated approach for larger ruptures in a particular zone to penetrate a neighbouring zone. This could be avoided if opting to delineate the zones around the partitions by hand. However, a degree of softness in the boundaries between zones is acceptable, since it is only the largest, and consequently rarest, ruptures that may penetrate the boundary. The variation in χ^2 with increasing K for the rupture catalogue is seen in Figure 4.14.

Perhaps the most interesting contrast between the rupture partition and the point partition is that there appears to be a clear "global" minimum when either the full or Poisson declustered catalogues are used without zones. In both instances it is the K = 29 and K = 30 partitions that produce the lowest χ^2 values. Since there appears to be a trend of increasing χ^2 for values of K above these values, it would appear that this is a minimum within the specified range of K being tested, rather than necessarily an inverse correlation between K and χ^2. This trend is not borne out when zones are used. Nevertheless, K = 29 appears as a "good" partition even when zones are used, suggesting that it is possibly the partition that is most able to reproduce the observed seismic hazard when using either a zoned or non-zoned approach.

As has been shown from the seismic hazard using the point source data, there is considerably more variability when using zones, compared to the non-zoned approach. Generally, the χ^2 values are higher when zones are used, the only exception in this instance being K = 2 when the Poisson declustered catalogue is used. Alongside the K = 2 and K = 29 partitions, several other partitions emerge as "better" than others: K = 27 and K = 32. There are also several partitions for which the full and declustered catalogues produce very different values of χ^2. For instance, K = 8 and K = 9 both produce low values of χ^2 when the declustered catalogue is used, but very high values when the full catalogue is

used (K = 9 being the maxima on this occasion). The same pattern can be seen in K = 22 and K = 35.

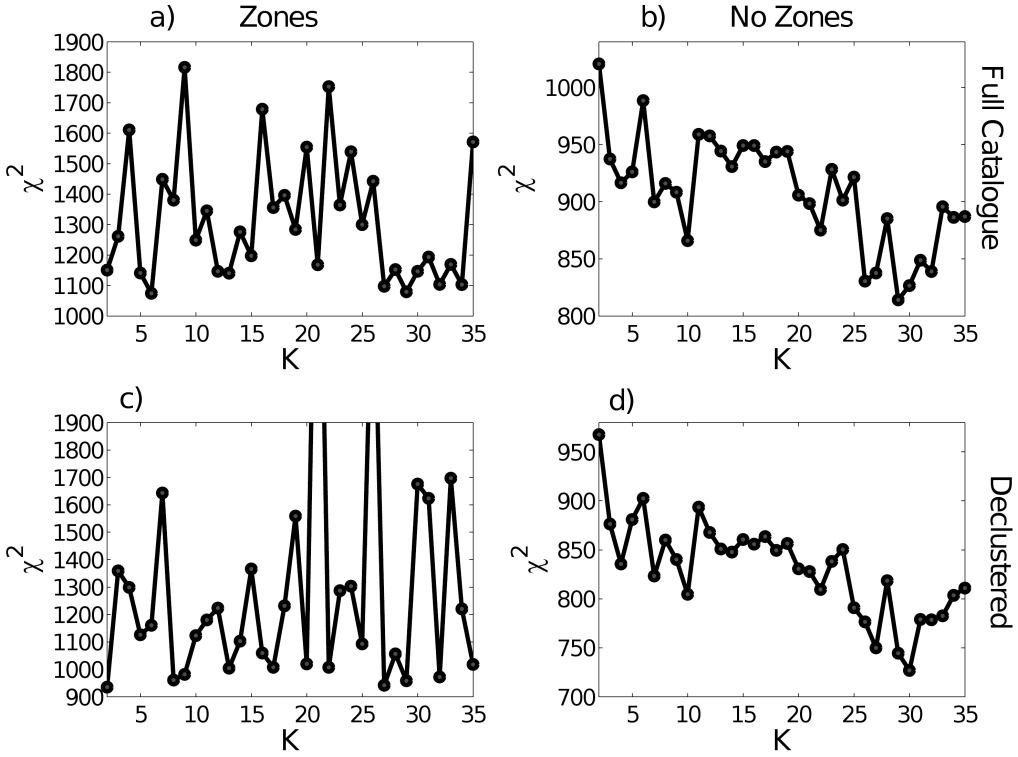

Figure 4.14: Variation in χ^2 values with increasing K for the Line K-means algorithm. The rupture catalogue is the same and the reference earthquake catalogue is the Aegean 20th century catalogue of shallow earthquakes ($M_W \geq 5.2$), a) Full catalogue with uniform zones, b) Full catalogue without zones, c) Declustered catalogues with uniform zones and d) Declustered catalogue without zones

4.6.5 The Optimum Number of Zones?

What has been seen in this section and the previous section is that, although some values of K stand out as producing consistently "better" fits than others, there is no single value of K that is identified as the optimum partition, via algorithmic methods alone. Whilst this would have been a desirable outcome, it is unsurprising that a definitive K has not been recognised. When simply using the validity indices, the number of clusters identified is influenced by the separability of the clusters. Conversely, the context-specific stochastic seismic hazard analysis method incorporates many more variables than those relating specifically to the spatial distribution of earthquakes or ruptures. Consequently the variation in χ^2 is greater and distinction in fit between partitions produced by varying values of K is harder.

These results clearly present an argument for selecting particular partitions for use in PSHA, but the issue of delineating a source model is not yet resolved. There is precedent in previous seismic hazard analyses to moderate different source models using a logic

tree. In circumstances where the seismicity may be described by several different models, this could be integrated into the epistemic uncertainty of the analysis. The weighting of each model with the logic tree may yet again be left to expert judgement, or may be determined mathematically by the strength of the χ^2 fit (or of the validity indices).

From a theoretical perspective, the method of identifying the optimum K, within the specified range, using stochastic seismic hazard analysis is essentially a one-dimensional inverse problem. The observed seismicity in a 100 year period is used here and compared with the synthetic seismicity of a large number of 100 year periods. The Monte Carlo seismic hazard calculations, therefore, form the forward calculation.

From a seismological perspective, the expression of seismic hazard analysis as an inverse problem is a novel and potentially challenging approach. Model selection in typical inverse problems is often undertaken with the use of information criteria such as the Akaike Information Criterion (AIC):

$$AIC = 2K_P - 2\ln(L) \tag{4.18}$$

where K_P is the number of parameters and L the likelihood function; the corrected Akaike Information Criterion:

$$AICc = AIC + \frac{2K_P(K_P - 1)}{n - K_P - 1} \tag{4.19}$$

where n is the sample size; or the Bayesian Information criterion.

$$BIC = -2\ln(L) + K_P \ln(n) \tag{4.20}$$

Where model errors are independent and normally distributed (valid here since it is the logarithm of ground motion that is used as the fitting parameter), then the likelihood function can be approximated by RSS/n, where RSS is the residual sum of squares. A full treatise of model selection is beyond the scope of this discussion, and the reader is referred to Burnham and Anderson (2002) for further detail. The model that minimises the information criterion is often deemed the "best model".

Unlike typical modelling scenarios, this application is limited by a large number of parameters for each model. Direct comparison of K and AIC or BIC, in the manner one might expect for other applications, is misleading. The implementation of stochastic seismic hazard analysis as a forward problem means that there are many more parameters to consider than simply K. Consider a partition of 10 clusters. It may be tempting to suggest that the number of parameters is 10. However, the seismic hazard analysis is a function of many more parameters for each cluster (a, σa, b, σb, M_{MAX}). Assuming each of these parameters to be independent then K would rise to 50. Since these parameters are not independent, and seismic hazard itself is a nonlinear function of these parameters, the theory of information criteria as applied to stochastic seismic hazard analysis begins to

break down.

Model selection via information criteria is underpinned by Occam's razor, "plurality ought never be posit without necessity". Under this assumption, the addition of more parameters, in this example higher K, must be rewarded by a substantial improvement in fit if it is to be justified. This makes parsimony a critical objective of model selection. "Inference under models with too few parameters can be biased, whilst with models having too many parameters there may be poorer precision or identification of effects that are spurious (Burnham and Anderson, 2004)". Figures 4.13 and 4.14 indicate that there is considerable similarity in χ^2 between models with "better" fits and those with poorer fits. When information criteria are applied, the penalty for the increase in the number of parameters far outweighs the small improvements in fit. Consequently AIC and BIC will likely identify the lowest K as the best model, and may imply even that K = 1 is the preferred model.

Ultimately, this section has illustrated the need to assess the partitions using judgments beyond the purely quantitative. In the next sections particular partitions, including those that have suggested good fits from the quantitative criteria, will be scrutinised with respect to the existing knowledge of seismotectonics, developed in Chapter 3 and Appendix A.

4.7 Appraisal of the earthquake (point source) partitions

From the previous sections it has been shown that identification of a single optimum partition via validity indices and stochastic seismic hazard analysis has not been possible. Analysis shall be made of those partitions that perform better from a purely cluster analysis perspective (i.e. those that were identified as "better" partitions by the validity indices): K = 4, K = 8, K = 25 and K = 36. Assessment is undertaken for a sample of the possible partitions for the zoned analysis: K = 5, 10, 20, 30, 40 and 50. Comparison is made with the (Papaioannou and Papazachos, 2000) model.

The K = 4 partition (Figure 4.15) essentially partitions the Aegean almost in terms of compass direction. The "North" zone covers the Adriatic coast and south Balkan region, whilst the "South" zone is limited to the central and Eastern Hellenic arc, encompassing most of the seismicity attributed to slab interface along the Hellenic subduction zone. The "West" zone covers most of central and Southern Greece and the Ionian Islands, whilst the "East" zone is mostly limited to Western Turkey. This partition represents a crude, but surprisingly consistent, delineation of some of the major tectonic features in the Aegean region. Most of the thrust events associated with subduction are separated from the extensional events of central Greece and Western Turkey. Equally, distinction is made between the high seismicity extensional region of central Greece and the lower seismicity extensional region of northern Greece and the southern Balkans. Some com-

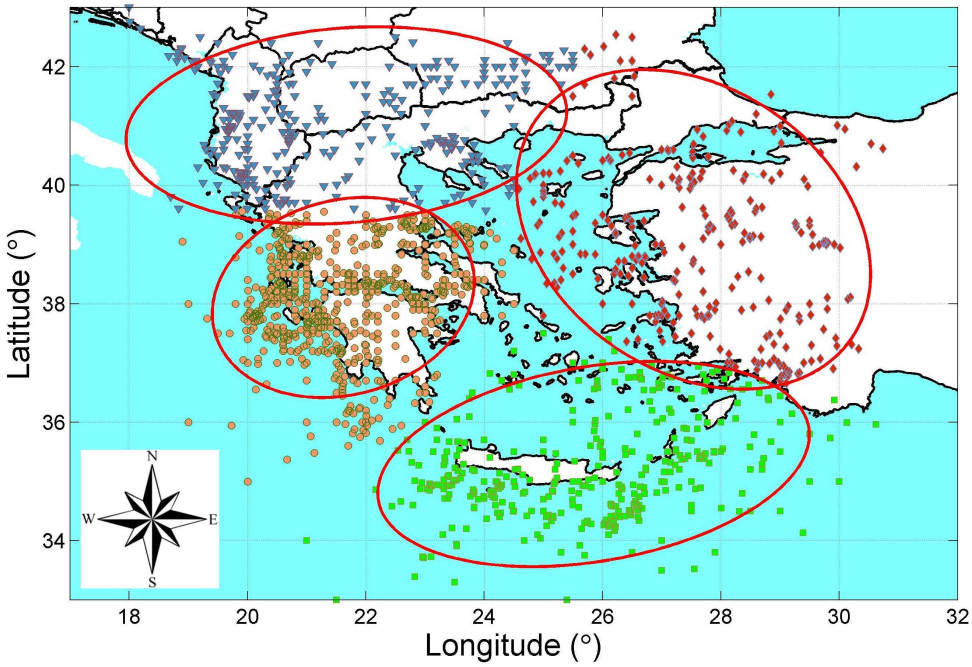

Figure 4.15: K = 4 partition of the declustered 20th century catalogue (subset 2)

parison can be made between the K = 4 partition and the McKenzie (1978) tectonic model, which broadly distinguishes between these regions (Aegean, Anatolia and Balkans) along similar divisions. Obviously a 4-cluster partition is unable to model all the variation in seismicity observed, but there are consistencies. For example the "East" zone terminates at much the same point in the Aegean Sea as the tract of the NAF. This therefore encompasses the strike-slip regime of the NAF and the extensional seismicity in southern Turkey. Perhaps the most erroneous division in this model is that of a small number of earthquakes around the Bulgaria-Turkey border region, which are incorporated into the "East" region, when they should perhaps be allocated to the "North" zone. This would place the boundary in the more logical place of the low seismicity area of northwestern Turkey. Similarly the "West" zone encroaches heavily into the Hellenic arc, thus mixing extensional, strike-slip and thrust faulting. Also the "North" zone contains both the thrust faulting of the orogenic collision along the Adriatic coast as well as the extensional faulting in the south Balkans.

By doubling the number of clusters from the K = 4 model, some of the inconsistencies begin to be resolved in the K = 8 model (Figure 4.16). The "North" zone in the 4-cluster model is now split into two zones: the first containing mostly thrust faulting along the Adriatic coast, the second containing extensional faulting earthquakes in Macedonia and southern Bulgaria, and logically terminating in the relatively aseismic Bulgaria-Turkey border region. Central and southern Greece is now split into three clusters: the Ionian Islands/Gulf of Patras, the Gulf of Corinth and the southern Peloponnese/West Hellenic arc. This begins to separate the extensional faulting of central Greece, from the strike-slip

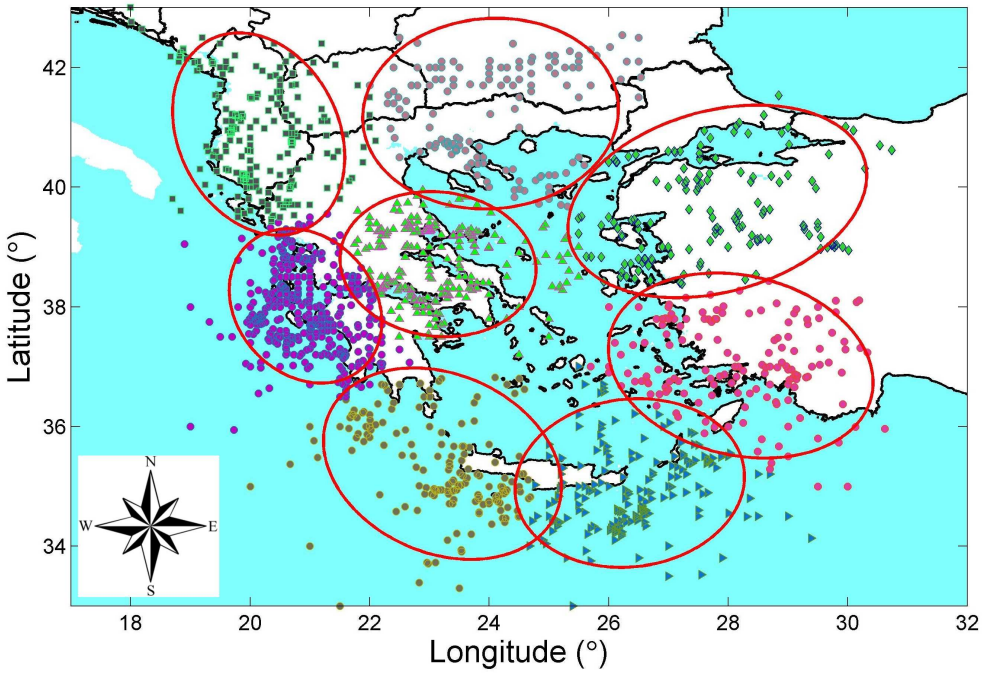

Figure 4.16: K = 8 partition of the declustered 20th century catalogue (subset 2)

regime found in the Ionian Islands, and also from the thrust faulting found in offshore in southern Greece. Western Turkey has now been divided into two clusters: the northwest, which incorporates most seismicity attributable to the western end of the NAF as well as some of the extensional seismicity further south, and the southwest, which incorporates the remainder of the extensional seismicity in southwest Turkey as well as the eastern section of the Hellenic arc. The Hellenic arc is now split into three sections, roughly corresponding to the change in strike of the plate margin from west to east; clearly visible in existing source models.

The K = 8 partition presents some interesting divisions, which are not necessarily in keeping with the observed differences in seismotectonics across the Aegean. This includes the extension of the south Balkan zone across a region of inactivity, thus incorporating seismicity from the NAF into the southern Balkan region. Other conflict is the mixture of thrust and extensional faulting in the eastern Hellenic/southwest Turkey zone. One conflict for which there is an argument both for and against is found in the Ionian Islands/Gulf of Patras region. Here the Ionian zone terminates in the western Gulf of Corinth, thus splitting an area of seismicity, rather than in the more obvious low seismicity region in the Gulf of Patras. The difference in observed deformation across the Gulf of Corinth (Clarke *et al.*, 1997) may provide a basis for delineating a zone boundary at the location that it has.

There is clearly an increase in resolution for the K = 25 model (Figure 4.17) as the clusters begin to become more compact. Particular divisions evident in this model that may be broadly consistent with differences in seismotectonics can be found in central Greece.

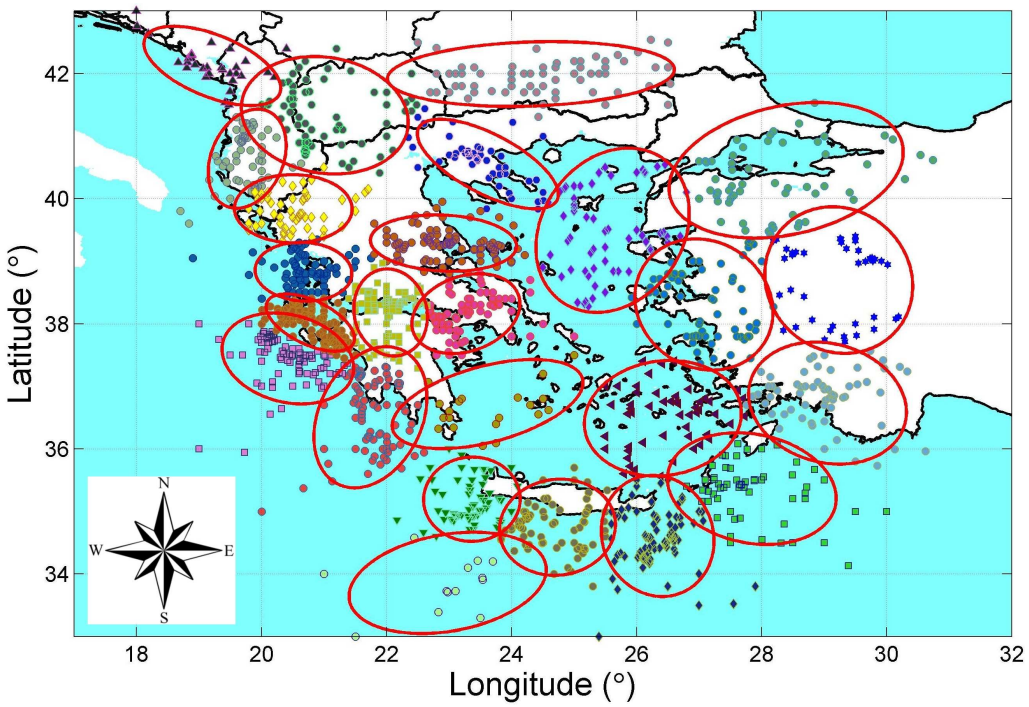

Figure 4.17: K = 25 partition of the declustered 20th century catalogue (subset 2)

Here, seismicity is described by four clusters: western Gulf of Corinth, eastern Corinth and Attica, Pagasetic Gulf and western Thessaly. These divisions reflect differences in observed deformation and seismicity in central Greece. Similarly, the NAF is split into several more compact clusters that follow its extension into the North Aegean Sea. More divisions are made in western Turkey, which are not dissimilar to those found in several earlier source models (PZ1990, PP2000). Where there are still some inconsistencies with the observed seismotectonics is in the Ionian islands and the western Hellenic arc. In these regions there are perhaps more clusters than is appropriate given the observed seismotectonics. Along the Hellenic arc the clusters are more compact, thus following more closely the trace of the plate margin, without extending as far as the regions of normal faulting in the back-arc region. In contrast, the offshore margin is described by six clusters, running east to west, which would suggest a disparity in the seismotectonics along the arc not really visible from other sources of seismotectonic information.

The K = 36 partition (Figure 4.18) demonstrates yet another increase in resolution, with many more compact clusters. One of the most interesting results of this model is the partitioning of the low seismicity area of the western Libyan Sea into a single offshore cluster. Separating these events, which pose little threat in terms of seismic risk in Greece, from the remainder of the arc means that the clusters follow the trace of the plate boundary more closely. There is further partitioning of seismicity in central and southern Greece. The Gulf of Corinth region is partitioned into three zones: Western Corinth Gulf, Eastern Corinth Gulf and Attica. However, the partition in the Thessaly regions remains much the same as in the K = 25 partition. Other similarities in the 25- and 36- cluster partition can be found in the Northern Greece/South Balkan region and in the Northern Sea

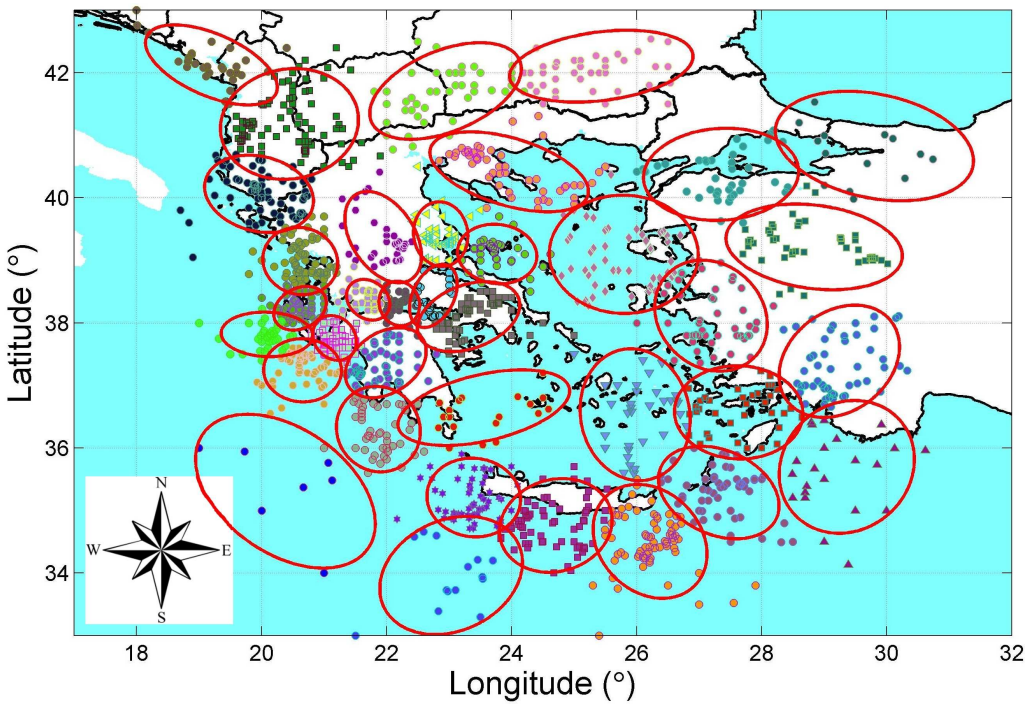

Figure 4.18: K = 36 partition of the declustered 20th century catalogue (subset 2)

of Marmara. One of the more notable differences between the two models is in the Ionian Islands. In the 36-cluster partition, the seismicity Ionian islands is described by five clusters. Given the high seismicity of the region, it is unsurprising that the K-means algorithm will sub-divide the region so extensively. Given the observed seismotectonics and faulting in the Ionian Islands (Appendix A), there is good reason to question the necessity of so many zones. Even in this region, the partition is actually not dissimilar to that of PP2000.

Another area of concern is in the western Marmara region. The K = 25 model had delineated a compact cluster of seismicity here, which contains many of the major strike-slip events found along this section of the NAF. However, in the K = 36 partition, the cluster now incorporates a cluster of normal faulting events further south. In terms of the seismic hazard analysis, this will have the impact of smoothing the hazard out over a larger area, which may lead to underestimation of seismic hazard in the western Sea of Marmara.

Once again, in the K = 36 partition there has been a further increase in the number of clusters into which the Hellenic arc is being partitioned. As suggested previously, it is unclear that there is a sufficient degree of seismotectonic variability along the arc to justify the number of clusters. Nevertheless, two sets of clusters are becoming increasingly apparent. One set follows the trace of the plate margin along the convex side of the arc, characterised mostly by low-angle thrust faulting. Another set follows the concave side of the arc in the lower seismicity south Aegean sea, where there is a greater degree of arc-parallel strike-slip and extensional faulting (Benetatos *et al.*, 2004). The CTEZ is also

now becoming visible in the form of two clusters that traverse its length.

These illustrations clearly demonstrate that as the number of clusters increases, smaller scale variability in the seismicity becomes more apparent in the partitions. In some locations this increase in variability is consistent with knowledge of the seismotectonics, often corresponding to abrupt changes in tectonic regime. Elsewhere it could be argued that the seismicity is over-partitioned, with regions such as the Ionian Islands being divided into separate, if not well-separated, clusters in spite of the consistent seismotectonic regime. This illustrates quite effectively the issue of parsimony associated with the theory of model selection, discussed in section 4.5.6. The occurrence of over-partitioned regions in the K = 36 model compared to the K = 25 model would suggest that from a seismotectonic perspective, an "optimum" partition would be found in between these two values, possibly closer to K = 25.

Recognising that an optimum partition, robust to changes in the input catalogue, has not been identified Figure 4.19 demonstrates the partitions for six different values of K: K = 5, 10, 20, 30, 40, 50. For ease of identification, ellipses are placed around the clusters. These ellipses are not illustrative of a zone and are not necessarily aligned to the strike of major faults in the area. Nevertheless, that the alignment of the major axis of the ellipse does sometimes correspond to the strike of major faults is not always coincidental.

The K = 5 model in Figure 4.19a bears a strong resemblance to the K = 4 model shown in Figure 4.15. The addition of another cluster has had the effect of reducing the area encompassed by each cluster. The greatest difference is found in western Turkey, where the "East" zone in the K = 4 model has been split most effectively into a northwest Turkey and southwest Turkey zone. Other changes include a compaction of the Hellenic cluster so that it is now centred on Crete, which represents an expansion of the southwest Turkey zone across the Dodecanese islands. The "West" cluster in the K = 4 model remains almost entirely unchanged in the K = 5 model, whilst the "North" zone now does not extend as far east into Bulgaria. As with the 4-cluster partition, the K = 5 partition has divided seismicity among the more major tectonic provinces of the Aegean. However, the mixture of compressional plate margin and mixed compression/extension intra-plate seismicity within the "North" zone is in conflict with many existing seismogenic source models. The same may be said for the Central and Southern Greece cluster, which mixes strike-slip, normal and thrust faulting into one large cluster.

A similar comparison in partitions can be made between K = 8 (Figure 4.16) and K = 10 (Figure 4.19b). Six of the ten clusters remain almost identical, with the only substantial changes occurring in the south and west of the Aegean region. In the K = 8 model, the region encompassing western Turkey and the central and eastern Hellenic arc was described by three clusters. In the K = 10 model this is now described by five clusters. The Hellenic arc is now split into two separate clusters, the boundary occurring in eastern Crete. This is in reasonably good agreement with the change in focal mechanism sug-

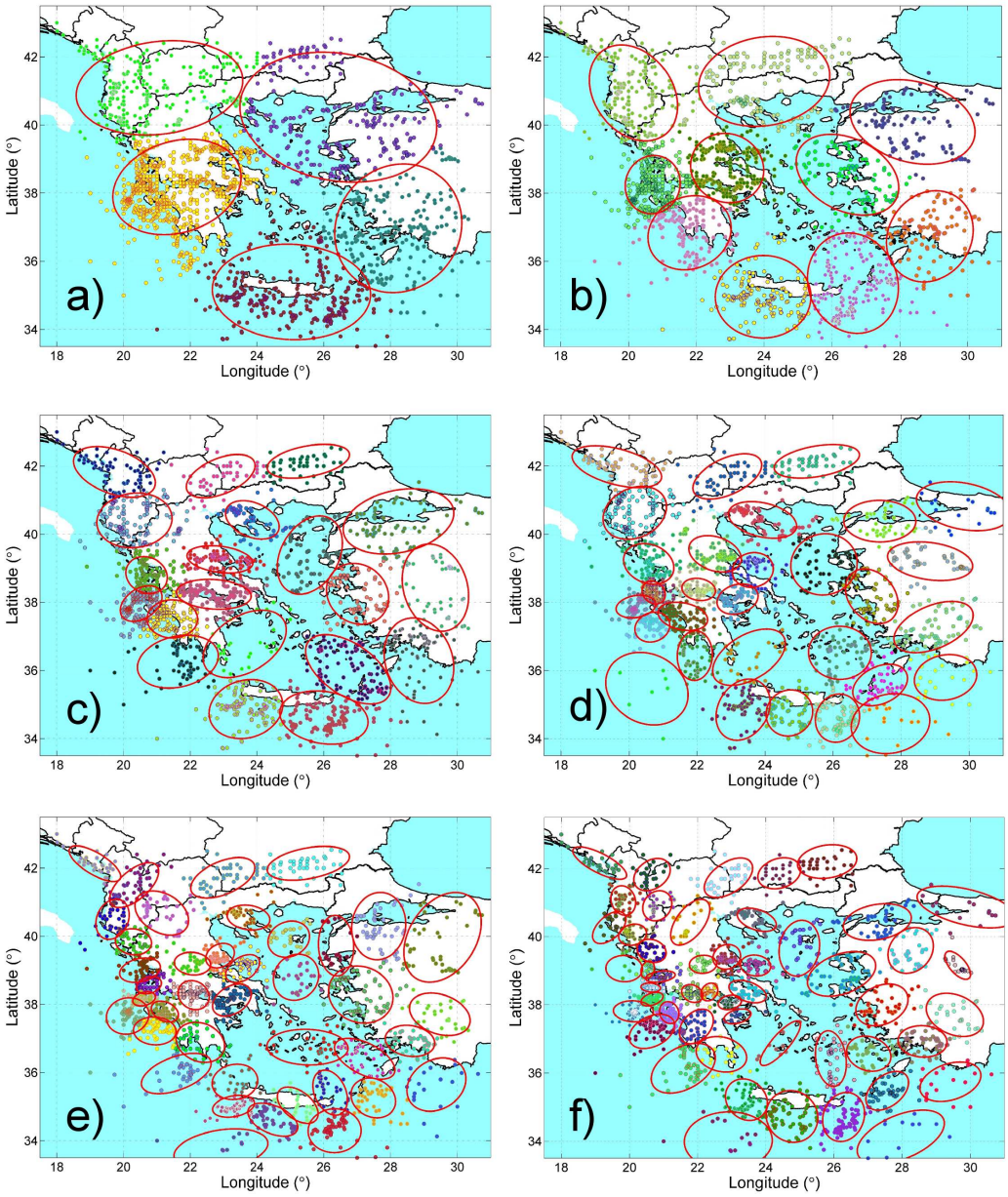

Figure 4.19: Partitions of the 20th Century Shallow (Depth $<$ 60 km) Aegean Catalogue ($M_W \geq 5.2$). a) K = 5, b) K = 10, c) K = 20, d) K = 30, e) K = 40 and f) K = 50. Red ellipses are used to distinguish between clusters and are not indicative of the size, shape or mechanism of the source zone

gested by (Benetatos *et al.*, 2004). The other substantial difference is the introduction of a new cluster, centred just off the western coast of Turkey. This zone broadly encompasses the mixed normal and strike-slip faulting associated with the southern branch of the NAF as it divides into two branches in western Turkey. The more active northern branch of the NAF is better described by a more compact and elongated cluster than in the K = 8 model.

The 20- and 30- cluster partitions offer perhaps the best compromise between seismic sources that are small enough to conform to local variation in seismotectonics, whilst still having a sufficient number of earthquakes in each cluster to determine b-value and M_{MAX}. In the 20 cluster model, there is considerable distinction between compressional, strike-slip and normal faulting regimes. Furthermore, the normal faults that typify much of mainland Greece and the southern Balkans are broken down into smaller clusters that appear well separated by regions of low seismicity. Good examples of this are the three E-W striking clusters that encompass the Gulf of Corinth, the Gulf of Volos and the Thessaloniki-Rentina Fault Zone (Tranos *et al.*, 2003) respectively. The trace of the North Anatolian fault is becoming visible by way of a series of clusters that closely follow the band of high seismic activity that extends from the eastern Sea of Marmara into the North Aegean Sea.

The 30-cluster partition makes further distinctions between groups of hypocentres, predominantly in the region of high seismicity around the Gulf of Corinth and Peloponnese. With more clusters being used, the K-means algorithm splits seismicity of the Gulf of Corinth into east and west sections, with the eastern section incorporating the normal faulting in the Parnitha region. A similar division is made in the Thessalia region of Greece, with the onshore seismicity of the Pagasitikos Gulf and surrounding region being separated from the offshore seismicity in the North Aegean Sea and northern coast of Evia island. Also, a distinction is made between seismicity in the eastern and western Sea of Marmara. Another useful result is that that the 30-cluster partition groups some of the isolated events found well-offshore in the Mediterranean into a widely-dispersed low-seismicity cluster. This particular cluster contains so few events, and is sufficiently far away from inhabited regions in the Aegean that it could reasonably be considered as having a trivial influence on engineering seismic hazard in the Aegean, or at least assigned as background seismicity. At the same time, it removes some of the outliers from clusters in the Hellenic arc, making them more tightly constrained and an improved representation of seismicity along the arc. In effect, this 30-cluster model could really be considered a 29-cluster model for the purposes of seismic hazard analysis. Comparison of the 30-cluster partition with that of the 25- and 36-cluster partitions begins to indicate a resolution of some of the modelling conflict that emerged in those models. This suggests that not only do the 29- and 30- cluster partitions perform strongly in the quantitative analysis of partitioning; they are also the most representative of the seismotectonic variation in the Aegean.

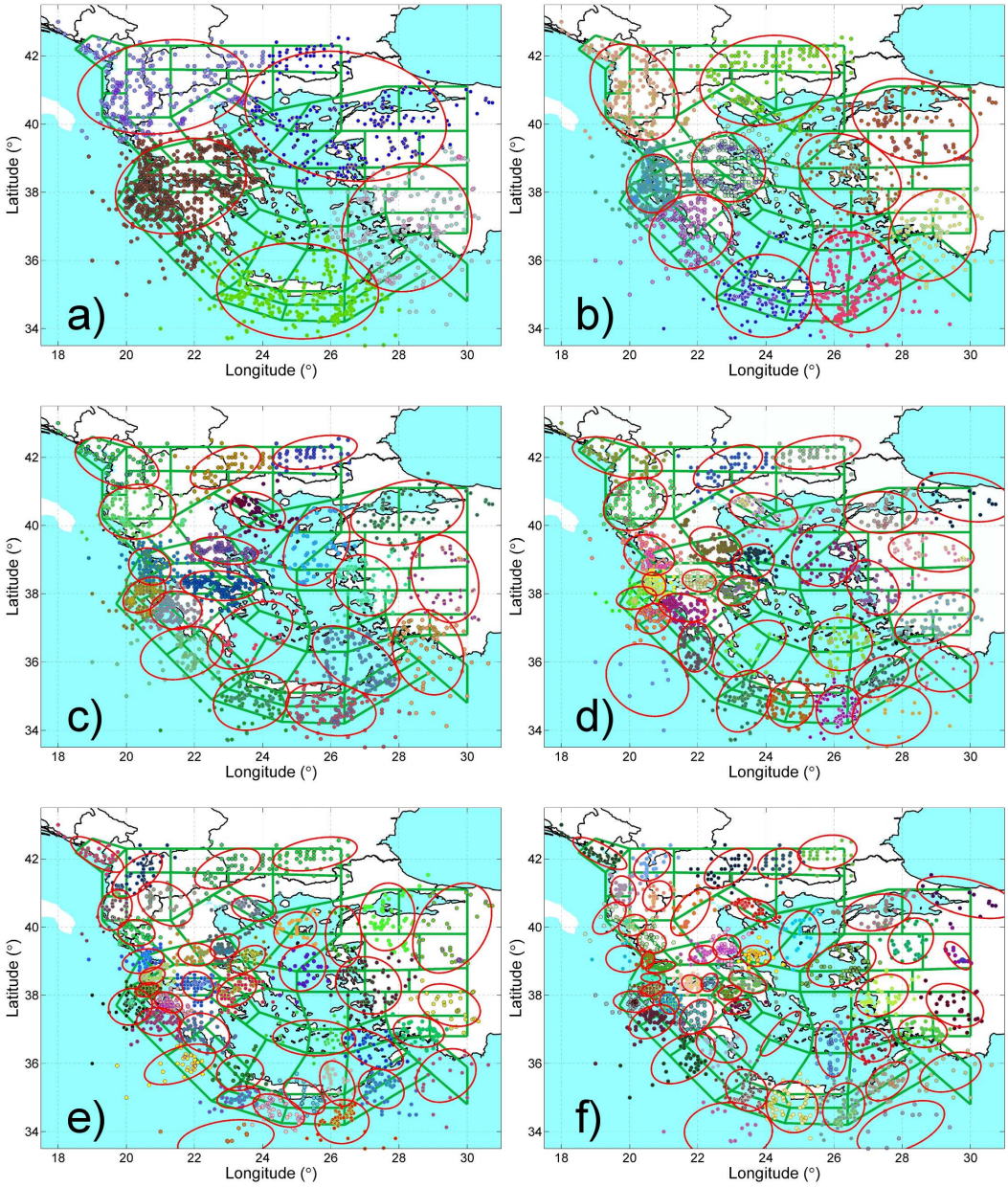

Figure 4.20: Comparison of the a) 5-cluster, b) 10-cluster, c) 20-cluster, d) 30-cluster, e) 40-cluster and f) 50-cluster partition with the Papaioannou and Papazachos (2000) source model (marked in dark Green). Ellipses (dashed lines) are used as markers and are not indicative of a source zone

The remaining models with higher numbers of clusters (K = 40 and K = 50, Figure 4.19e and 4.19f, respectively) continue this trend of subdividing areas of high seismicity into smaller groups. In doing so, many sets of neighbouring clusters emerge, for which the seismotectonic properties are so similar that they may be indistinguishable when their uncertainties are taken into account. This is exacerbated by the fact that with fewer earthquakes in each cluster, the uncertainties on particular properties such as b-value, M_{MAX} and strike are going to be greater. By over-partitioning the hypocentres there is a greater risk that the delineation of a source zone will be made based upon a transient feature of seismicity within the observed catalogue, even when a declustered catalogue is used. Although Poisson declustering suggests a time invariant catalogue with properties of stationarity, it will be the case with increased subdivision that features and manifestations with intrinsic return periods beyond the observed catalogue duration will be ill-determined. As the K = 50 model offers no robust improvement of fit over the 30-cluster model, then the 30-cluster model should be preferred. Exception to this approach might be made where evidence for a higher cluster model comes from information pertaining to the physical properties of faulting in the region not adequately captured by the seismicity. Despite the growing knowledge of Aegean tectonics, it is not believed that such an argument can be made in this example, and hence one would recommend use of the 30-cluster model rather than the 40- or 50-cluster model in a seismic hazard analysis, or, if several models are used, should be weighted accordingly.

These partitions are compared with the PP2000 source model (Figure 4.20), and the automated uniform zones shown in Figure 4.21. Obviously, it is difficult to draw comparison with the 67-zone PP2000 model when considering partitions with very few ($K \leq 20$) clusters. Some interesting features emerge in the 30-, 40- and 50-cluster models, which may present an argument for grouping some of the PP2000 zones. In the 30-cluster partition many of the clusters demarcated in the Ionian Islands and Central Greece broadly correspond to the zone boundaries of the PP2000 model. The cluster encompassing the western Gulf of Corinth divides seismicity in the same manner as a single zone that incorporating zones 42 (Patra), 43 (Aeghio) and 39 (Agrinio) of PP2000. Similarly the cluster of hypocentres in the eastern Corinth and Parnitha region encompasses zones 41 (Thebes), 44 (Corinth), 45 (Methana) and 51 (S. Euboikos) of PP2000. These particular similarities persist in the 40-cluster model. There are still many discrepancies between the models, particularly in the eastern Aegean region. Furthermore, none of the cluster models are particularly adequate in capturing the narrow band of high seismicity that runs from the Cyclades islands to Western Turkey. In this region, the PP2000 model may be more suitable.

Figure 4.21 shows the tessellated zones that the partitions in Figure 4.20 produce. It is at this point that some of the similarities between the partitions and the PP2000 model begin to disappear. Clearly the tessellated zones will not necessarily reflect the shape of the major fault segments in the manner that some of the zones in the PP2000 model do. Nevertheless, the tessellated zones are reflecting the changes in seismicity, between

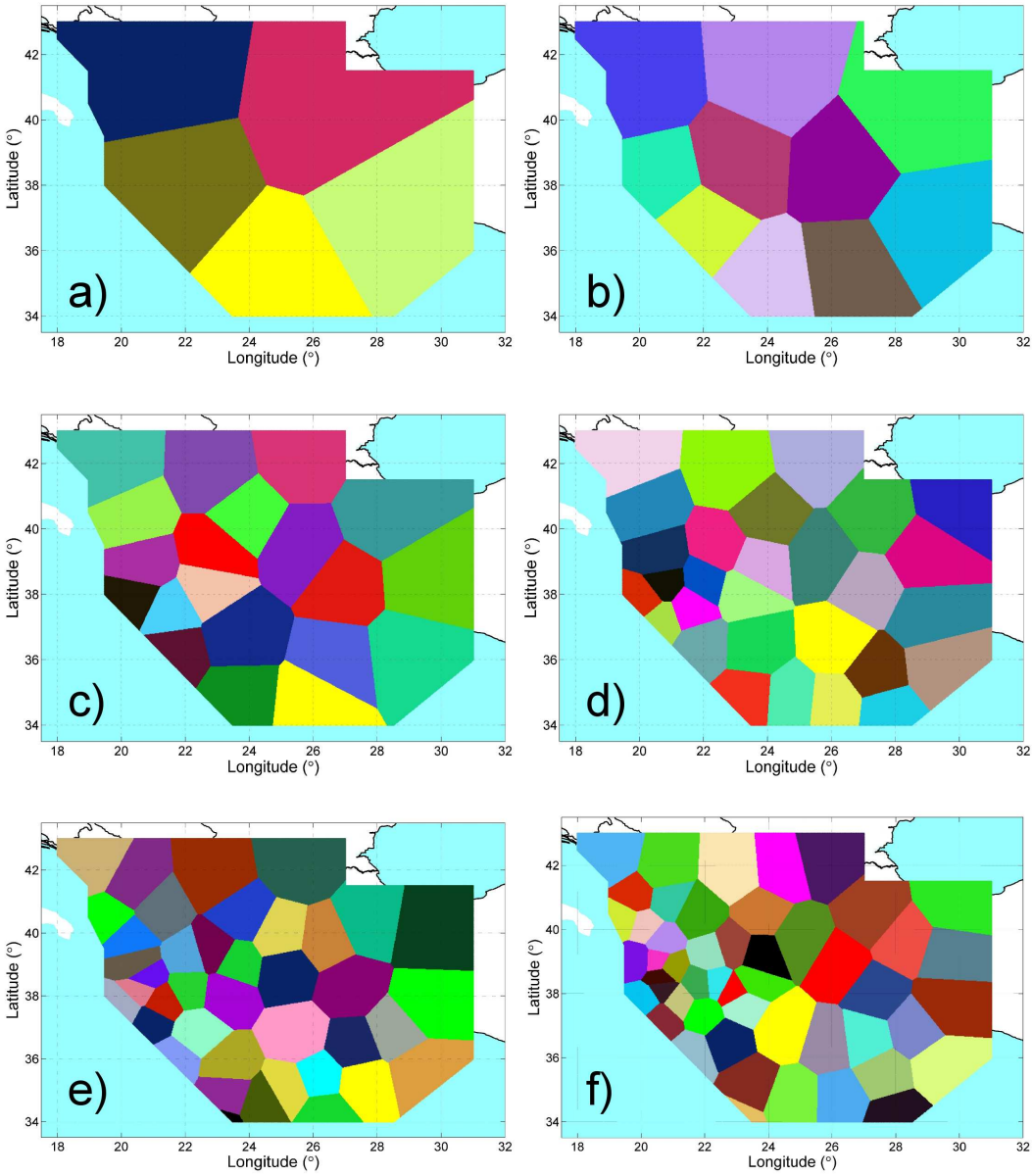

Figure 4.21: Uniform zones created by partitioning around the centroids of the shallow Aegean 20th century earthquake catalogue: a) 5-cluster, b) 10-cluster, c) 20-cluster, d) 30-cluster, e) 40-cluster and f) 50-cluster

highly active and relatively aseismic areas. It is not surprising that source models produced by eye will tend to fit regular shapes. The exact delineation of a zone boundary is perhaps the most subjective part of zonation, since it can be influenced both consciously and sub-consciously by the desire for regularity in the zones. Using automated tessellation in the manner described here, the zone boundaries only extend as far as the area of influence of the cluster.

4.8 Appraisal of the rupture partitions

As with the point-source partitioning, quantitative analysis of the partitions with respect to K has not consistently identified an optimum partition. Consequently it is necessary to inspect some of the "better" partitions more thoroughly in the context of Aegean seismotectonics. The following rupture partitions will be scrutinised in further detail: K = 10, K = 13, K = 15, K = 22, K = 27, K= 29 and K = 30 (Figures 4.22a to 4.22g, respectively).

Much like the hypocentre partitions, the rupture clusters when K is low broadly demarcate some of the larger tectonic features in the Aegean. In the K = 10 example, although the locations of the clusters are far from identical, there are many similarities between the point partition and the rupture partition. In Greece and the Hellenic arc the two clustering schemes generally separate out the same regions: the Adriatic coast, the Ionian Islands and Gulf of Patras, The Gulf of Corinth and the Attica-Thessaly region, southern Greece and eastern Crete. The differences occur mostly in northeast Greece and the eastern Hellenic arc. When ruptures are used, northeastern Greece is partitioned into two clusters rather than one, whilst the Dodecanese and western Turkey are amalgamated into one cluster here. Also, the historical seismic gap between the 1912 rupture in the western Sea of Marmara, and several historical ruptures in the eastern Sea of Marmara, means that the northern branch of the NAF is partitioned into two clusters, with the westernmost ruptures being clustered with extensional faults in southern Bulgaria.

Some agreement between the clusters and faulting types is visible now in the 10-cluster rupture partition. For example the transition from strike-slip faulting along the Cephalonia transform to thrust faulting in the western Hellenic arc is now more clearly demarcated. Similarly, the more active extensional region of the Gulfs of Corinth and Evia are separate from the Northern Greece/Southern Balkans region, which displays intraplate seismic behaviour consistent with a lower strain rate.

The 13- and 15-cluster rupture partitions resolve some of the conflicts in within-cluster fault type that had emerged in the 10-cluster partition. In the 15-cluster partition the trace of the NAF is becoming more visible in the form of three compact clusters, encompassing eastern Marmara, the Gallipoli peninsula and the North Aegean Sea respectively. A further change is that a new cluster emerges, in which the NW-SE striking thrust faults in Montenegro reside. Perhaps the most significant difference between the 13- and 15-

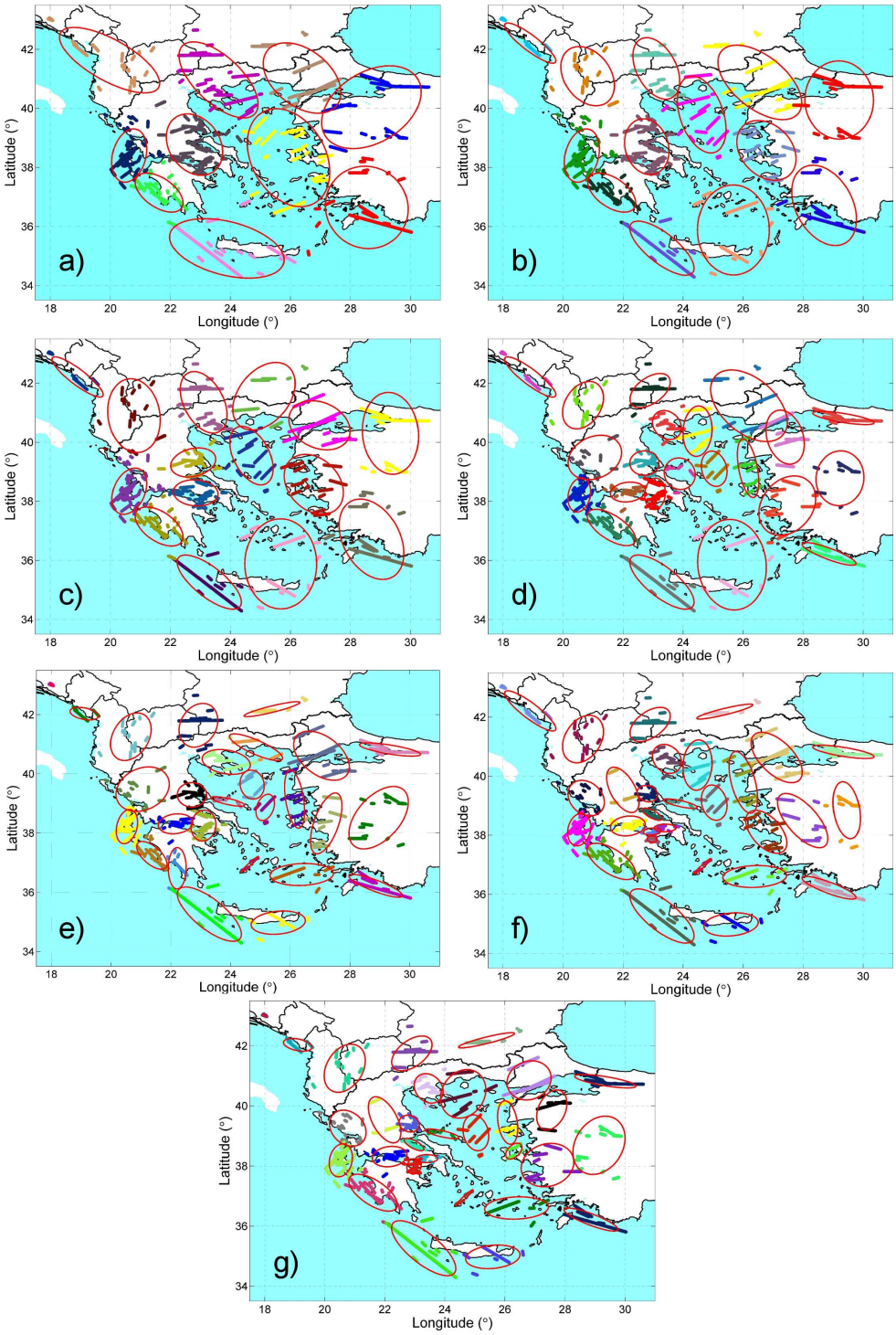

Figure 4.22: Partitions of the modified catalogue of known ruptures. a) K = 10, b) K = 13, c) K = 15, d) K = 22, e) K = 27, f) K = 29 and g) K = 30

cluster partitions is the division of central Greece into two clusters: Thessaly and eastern Corinth. This division is in good agreement with the variation in strain rate and seismic activity, with eastern Corinth displaying greater activity and strain than the Thessaly region.

Most of the new clusters that can be found in the K = 22 model are in central Greece and the Aegean Sea. The Gulf of Corinth is now partitioned into an eastern cluster and a western Cluster. A new small cluster is introduced north of Evia Island, which may represent a transition in seismicity from the oblique faulting in the North Aegean sea to the E-W striking extensional faults in central Greece. It is interesting that a band of small clusters runs from the Ionian Sea all the way across the Aegean to the eastern Sea of Marmara. Some clusters within this band follow the trace of the southern branch of the NAF, which would implicitly suggest a tectonic link between NAF and the strike-slip faulting in the Ionian Islands. Such a link has been posited (Goldsworthy *et al.*, 2002; Reilinger *et al.*, 2006) but there has been little in the way of tectonic evidence to support this. One suspects that the emergence of a link in the 22-cluster model is coincidental, as it begins to disappear in the higher K partitions. Another interesting result to note is how robust the clusters along the plate margin have been to higher K. Where new clusters have been inserted they are mostly in the central Greece and central Aegean region. This may perhaps suggest that the plate margin is well defined in the rupture catalogue, implying stability in the zonation process in active plate boundaries.

The remaining models (K = 27, K = 29 and K = 30) begin to resolve many of the concerns of previous models, which arose due to conflicting fault types being partitioned into the same cluster. In all three models the broad trace of the North Anatolian Fault is emerging as a series of smaller clusters that run east to west across the Sea of Marmara, and then turning southwest into the North Aegean Sea. In doing this the K-means method has segmented the NAF along its length making it harder to detect the traces of the northern and southern branches. This is representative of a dilemma that exists in zoning this part of the Aegean. Some source models (PP1990, PP2000) have chosen to segment the western end of the NAF, whilst others (EK1999) prefer to keep the NAF as a single coherent structure. That the K-means method segments the fault is simply a manifestation of the discrete distribution of fault segments from the Papazachos *et al.* (1999) catalogue. From a seismotectonic perspective, the partitioning of the NAF into northern and southern branches, and the increasing extensional component of slip as it enters the North Aegean basin, would strongly suggest that modelling the entire fault as a single entity is not necessarily the most appropriate approach.

The 29- and 30- cluster models generally agree along the length of the plate margin, from the Adriatic Sea to the Dodecanese Islands. The principal difference being that the 30-cluster model splits the NW to SE striking thrust faulting along the Montenegro coastline into two clusters (one of which is a singleton), instead of one. For the purposes of considering seismic hazard in the Aegean region, this distinction may not be that relevant.

Perhaps the most interesting distinction between these two models arises in the eastern Corinth region. The Thessaly region is consistently divided into two clusters, one containing the high seismicity around the Pagasitikos Gulf, the other the lower seismicity extending from the Kardista basin northwards to include the 1995 Kozani-Grevena earthquake rupture. Furthermore, both models clearly identify the Atalandi fault zone as one cluster, which is dominated by the ruptures associated with the two large earthquakes of April 1894 (M_W 6.7 and M_W 7.2). This disparity arises in the region around the city of Corinth, where surprisingly the 29-cluster models splits the ruptures into two clusters whilst the 30-cluster model keeps it as one. The same distinction between the western Gulf of Corinth and the Attica region remains.

In Figure 4.23 the higher K rupture partitions are compared with the PP2000 shallow seismogenic source model. Again, whilst there is clearly a disparity in the number of zones used, many similarities emerge too. In the 22-, 29- and 30- cluster models, the cluster over the Ionian Islands is a very close match to an amalgamation of zones 6 (Leukada), 7 (Cephalonia), 9 (Pylos) and 11 (Ionian Sea 1) in the PP2000 model. These four zones display similar seismic properties, which would lead to the suggestion that maybe the PP2000 model over-partitions the Ionian region. A similar match can be seen between the 30-cluster model and the PP2000 model central Greece, with zones 37 (Thessalia), 40 (Maliakos), 41 (Thebes) and 44 (Corinth) of PP2000 being well replicated by the rupture partition. There are many other areas, which whilst they may not display an exact match, identify similar boundaries between neighbouring zones for both the 29- and 30-cluster models and PP2000. The decision to partition the Sea of Marmara into two zones in the PP2000 model is supported by the rupture partitions, again due largely to the seismic gap in the central Marmara Sea. Similarly, the four zones spanning Albania and western Macedonia in PP2000 [2 (Dyrrachium), 3 (Avlona), 21 (Piskope) and 22 (Ochrida)] are all encompassed into one cluster in both the 29- and 30- cluster partitions.

An important question to consider is whether a database of 223 ruptures is sufficiently representative of the tectonics of the Aegean region as to form a reliable basis for zonation using cluster analysis. There are some regions of significant seismic activity that are poorly represented in the catalogue of known fault ruptures. Of these, perhaps the greatest mismatch between observed seismicity and known fault ruptures is found along the western coast of Northern Greece and Albania. Similar mismatches can be found around the Trikala province of central Greece and the areas of offshore seismic activity around the island of Karpathos and immediately to the north of Crete. Conversely, the rupture associated with the A.D. 365 Gortyna earthquake ($\approx M_W$ 8.3) dominates the Hellenic arc to the west of Crete; an area that more recently has experienced lower seismicity than along the rest of the arc. Given the considerable uncertainty associated with this earthquake's magnitude and location, and consequently with the rupture length, this may have an unduly large influence over the K-means partition in the Hellenic arc. It is certainly recommended that the K-means algorithm is repeated as more ruptures are discovered; be it in the course of an earthquake or by re-evaluation of historical events.

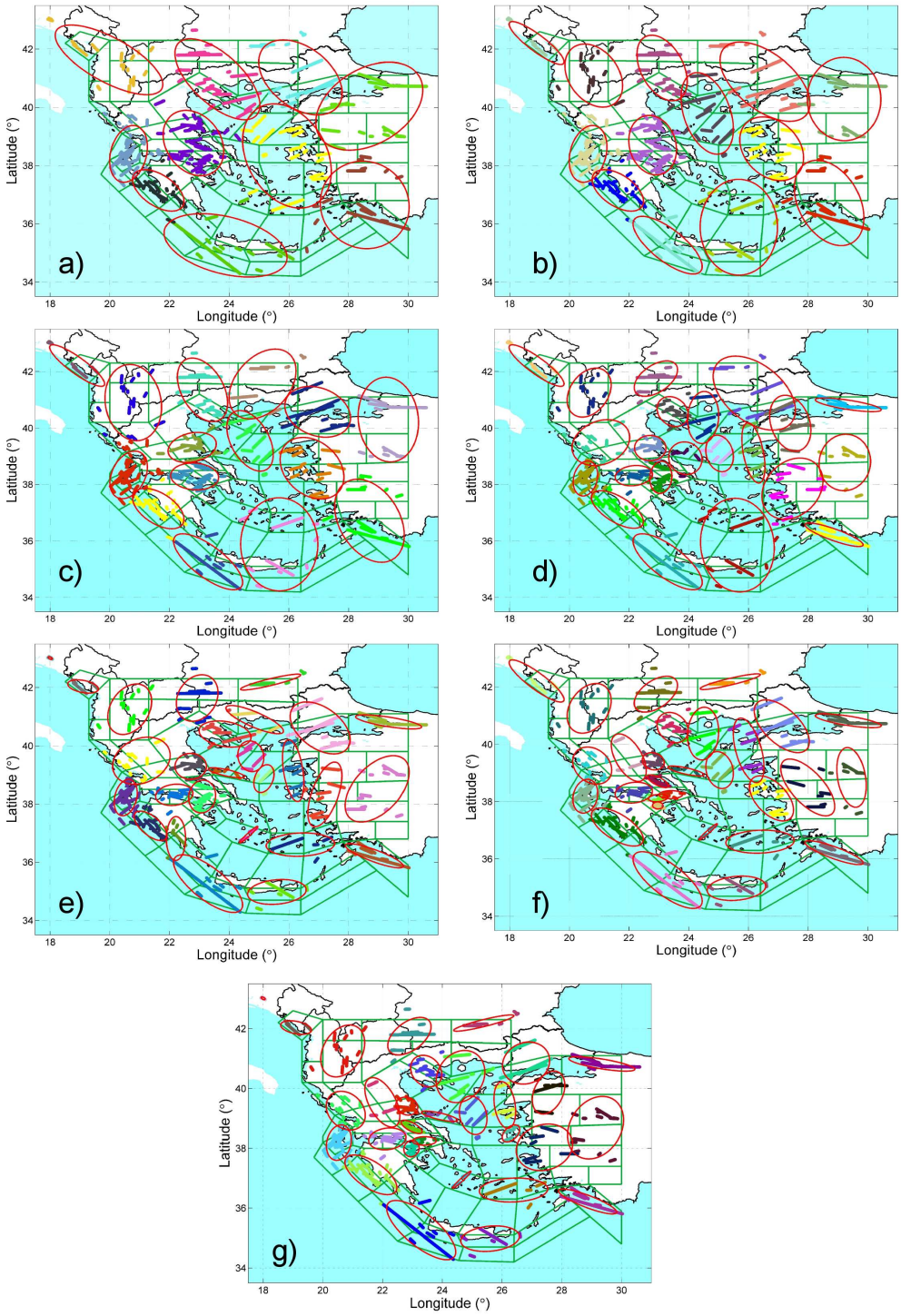

Figure 4.23: Comparison of the a) 15-cluster, b) 22-cluster and c) 29-cluster and d) 30-cluster partition with the source model of Papaioannou and Papazachos (2000) (marked in dark Green). Ellipses are used as markers and are not indicative of a source zone

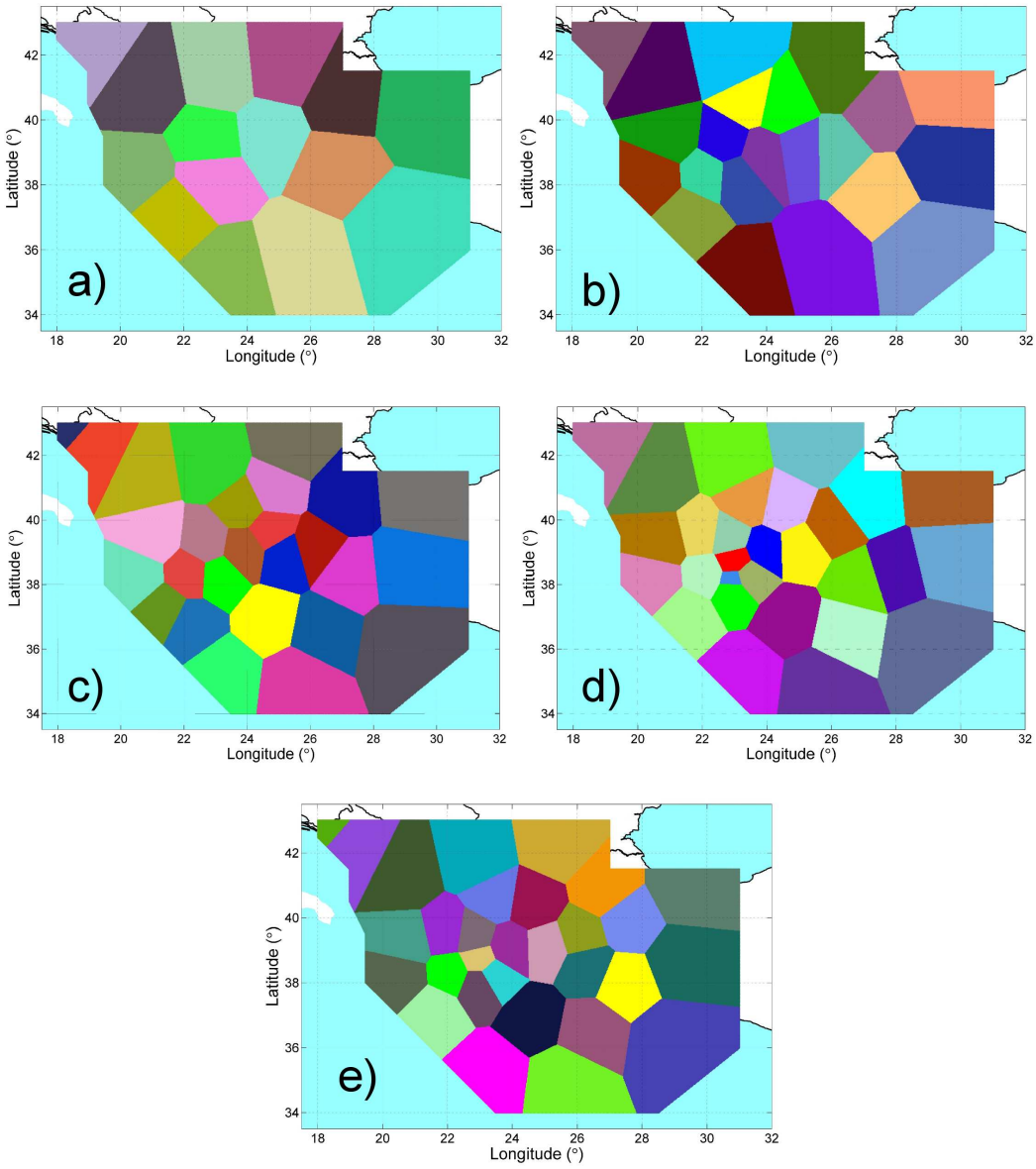

Figure 4.24: Uniform zones created by partitioning around the centroids of the rupture catalogue. a) 15-cluster, b) 22-cluster, c) 29-cluster, d) 30-cluster

The tessellated zones for the rupture partitions can be seen in Figure 4.24. As with the hypocentre partitions in Figure 4.21, there is not an especially great agreement between these models and the PP2000 model. There are still some similarities. In the Ionian Islands both the 29- and 30-cluster models contain a zone whose boundaries broadly mirror that of the aforementioned zones in the PP2000 model. Similarly, several tessellated zones along the NAF correspond closely to zones in the PP2000 model. The main differences are seen along the Hellenic arc, where PP2000 delineates a narrow group of rectangular zones along the plate margin, whereas here the plate margin manifests as a series of larger irregular polygons. This may have the impact of smoothing hazard along the plate margin, leading to underestimates of the likely ground motion.

4.9 K-means approach: Discussion and Future Directions

4.9.1 Appraisal of the K-means methodology

The source models presented demonstrate how the K-means algorithm can be used to delineate seismic sources, without necessarily introducing additional information about the seismotectonics of a region. They should be appraised in the context of the seismotectonics, especially where several models may appear to produce similar fits to the observed seismic hazard, as they do here. The motivation for applying this procedure is to institute a degree of consistency into the procedure of developing seismic source models. Furthermore, by implementing an algorithmic approach the zonations are developed (and analysed) with a greater degree of objectivity than would otherwise occur if delineating by eye.

The question therefore arises as to what degree has consistency and objectivity in source delineation been achieved by implementing the K-means methodology? It is important to recognise that the partitions presented arise from a stochastic process. This occurs not in the K-means algorithm itself but in the seeding of the initial centroids. The ensemble analyses identify an optimum partition that is not necessarily a global optimum for the specified K. It should be noted, that when ensembles were repeated or extended to include a greater number of trials, the differences in the partitions produced were often minor (with only a very small number of elements moving) or not apparent. When the total number of elements N in a dataset was small, such as in the rupture catalogue, there was an even greater degree of stability in the partitions.

It emerges that even if a global optimum partition is found for a particular data set and number of clusters, the optimum K is sensitive to changes in the data set [catalogue] when using the optimum partition searching methods presented. In the analyses presented here, one would refrain from identifying an optimum number of clusters using just the quantitative assessment methods shown. Only by analysing partitions in the context of existing knowledge of Aegean seismotectonics, has it been possible to justify particular

partitions suitable for extension to source models.

Clearly a degree of subjectivity remains in the source modelling process. Although quantitative objectivity was a motivation in the development of this methodology, it is perhaps an unachievable goal in isolation. To deputise the K-means algorithm to the extent that it has the same nuanced appreciation of earthquake behaviour as a working seismologist would be an inappropiate approach. K-means is an unsupervised algorithm designed to partition a set of data in such a manner as to minimise TWCSS, given an initial set of centroids. What the K-means algorithm does achieve is the refinement of a large (if not quite infinite) set of possible source models into a small number of possible delineations that can be compared both subjectively (by user analysis) and objectively (by the stochastic SHA procedure suggested here). It is also recommended that analysis such as those presented here should be repeated as more data become available, both in the earthquake catalogue and the rupture catalogue.

Where the data contains compact well-separated clusters, a consistent and global optimum may be found and may be robust. For the distribution of hypocenters and/or ruptures in the Aegean, this is not the case. Spatial clusters of earthquakes in the Aegean are usually not well-separated. Furthermore, earthquakes have also occurred in regions well away from clear tectonic margins and large-scale active fault structures. Their inclusion into the data set means that they are also subject to partition. This can result in isolated earthquakes, presumably with ill-determined physical features, being attributed to clusters in such a manner that may appear unrealistic to a person well-acquainted with the seismotectonics of the region in question. When attempting to delineate uniform source zones, this may have the impact of expanding such a zone over a much greater area than is appropriate to adequately model the seismic hazard in the region. In lower seismicity regions, however, it may be the case that seismicity is well distributed spatially, and hence attribution to an active seismogenic structure may not be possible. In these circumstances where information regarding the seismotectonics of a region may not be comprehensive, cluster analysis may prove a substantially more robust method of delineating zones with similar seismic properties.

4.9.2 Implementing these partitions in a seismic hazard analysis

The first question to address is which of the source models presented here should be implemented in the next stage of the seismic hazard analysis. Several partitions emerge as being possible candidates, and in particular the K = 29 and K = 30 models appear most consistent with the observed seismotectonics of the Aegean. The quantitative analysis would suggest that other models may be suitable, including: K = 22 and K = 27 (for the rupture catalogue) and K = 25 and K = 36 (for the point catalogue).

Perhaps the most obvious way to integrate several models into seismic hazard analysis

is via a logic tree, or further Monte Carlo methods (Cramer *et al.*, 1996; Smith, 2003). The model with the lowest χ^2 fit may be weighted strongly, whilst also alternative models with poorer fits are considered. However, a quantitative weighting scheme does not take into account the relation to the known Aegean seismotectonics that have been discussed here.

The use of tessellation as a method of delineating uniform seismic source zones around clusters is a relatively naïve procedure aimed at quantitative analysis of the partitions. These zones represent one of the possible delineations of seismic source models around a single partition. It may be the case that a particularly poorly fitting partition in the χ^2 sense may be improved by drawing the source zones around the clusters by inspection. This would allow for the inclusion of yet more knowledge, as reasonable estimates of the extent of uniformity in a cluster can be based on the judgement of an experienced seismologist, and not on an automated technique.

The question arises as to where these partitions fit within the field of existing source models in the Aegean region. Comparison has been made with the PP2000 source model, which is representative of the state of knowledge of seismicity and tectonics at the time of its production. It is not suggested that any of the cluster models here should be given preferential consideration over existing source models derived by other authors. Existing models of the Aegean can be tested alongside the source models presented here for comparison, or even for weighting. Since they have not been derived by the same method and not just derived from the observed seismicity distribution, quantitative comparison may be unfair.

For the purposes of the next stage of the seismic hazard analysis, the following tessellated source models will be used:

1. K = 29 (Rupture Catalogue): This is arguably the most consistently well-performing model when the stochastic seismic hazard analysis is used. The partition also performs well when analysed in the context of Aegean seismotectonics.

2. K = 30 (Rupture Catalogue and Point Catalogue): This model does not perform quite as consistently as the K = 29 model. It is, however, reasonably representative of the observed seismotectonics. The point source partition is also the only model for which a cluster is located so far offshore as to be deemed outside the overall source area. This cluster could be removed without any notable influence on seismic hazard onshore, essentially creating a 29-cluster model. This again gives a degree of credence to the use of the 29-cluster model as the optimum.

3. K = 27 (Rupture Catalogue): This is, again, another consistently well performing model in the stochastic seismic hazard analysis, with and without uniform zones.

4. K = 32 (Rupture Catalogue): This model performs reasonably well in the stochastic

hazard analysis, and is remarkably robust to the choice of input catalogue.

It can be seen from this selection that it is the rupture partitions give a generally better fit to both observed seismic hazard and seismotectonics in the Aegean region. These partitions tend to be more robust to the input catalogue used. By means of comparison with existing source models: PZ1990 contains 36 zones, EK1999 21 zones (for the Aegean region), PP2000 67 zones and JM2001 43 zones. The trend in χ^2 seen in Figure 4.14 for the non-zoned rupture data suggests an optimum number of clusters of K = 29 or K = 30. Given the issue of parsimony that has been raised previously, the K-means analysis here would suggest that source models in the range K = 25 to 36 may represent the optimum partition of the region. It would also imply that the PP2000 model is over-fitting the Aegean.

4.9.3 An alternative to tessellation

The use of the perpendicular bisector of neighbouring zones as a tool for delineation of zone boundaries is simple and effective, but does not necessarily take into consideration the shape of faults in each zone. When partitioning a set of evenly distributed point data this may not necessarily be an issue. When partitioning ruptures, it may be desirable to delineate zones in accordance with the orientation of rupturing. This can, of course, be done by manually using the partitions as a guide for zone shape. If opting to automate the process, the following method can be implemented as an alternative.

For each cluster, a linear set of discrete, evenly spaced points is defined along the length of all the ruptures in the cluster (a spacing of 200m is used here). The set of points is then smoothed across a grid of $0.1^\circ \times 0.1^\circ$ spacing, using the spatial seismicity smoothing method (Frankel *et al.*, 1996; Stirling *et al.*, 2002). The smoothed number of points in each cell (N_{si}) is calculated by:

$$N_{si} = \frac{\sum_j \left(N_j \exp\left[-d_j^2 / c_i^2 \right] \right)}{\sum_j \left(\exp\left[-d_j^2 / c_i^2 \right] \right)} \tag{4.21}$$

Where N_j is the number of rupture points in grid cell j, d_j is the distance between the centre of the current cell and the centre of cell j, c_i is the correlation distance (assumed here to be a constant 50 km). The correlation distance can be varied in accordance with the length of observed faulting within the cluster being considered if deemed necessary.

Once a grid of smoothed rupture points is defined a two dimensional Gaussian function is then fit to the points:

$$f(x,y) = Ae^{-\left(a(x-\bar{x})^2 + 2b(x-\bar{x})(y-\bar{y}) + c(y-\bar{y})^2\right)} \tag{4.22}$$

$$a = \frac{\cos^2\theta}{2\sigma_x^2} + \frac{\sin^2\theta}{2\sigma_y^2} \qquad b = -\frac{\sin 2\theta}{4\sigma_x^2} + \frac{\sin 2\theta}{4\sigma_y^2} \qquad c = \frac{\sin^2\theta}{2\sigma_x^2} + \frac{\cos^2\theta}{2\sigma_y^2} \tag{4.23}$$

Where $\bar{x}$ and $\bar{y}$ are the mean points, with standard deviations of σ_x and σ_y respectively; A the amplitude of the function and θ the angle of rotation relative to the longitude axis. This process is then repeated for all the clusters until a functional surface is formed across the region.

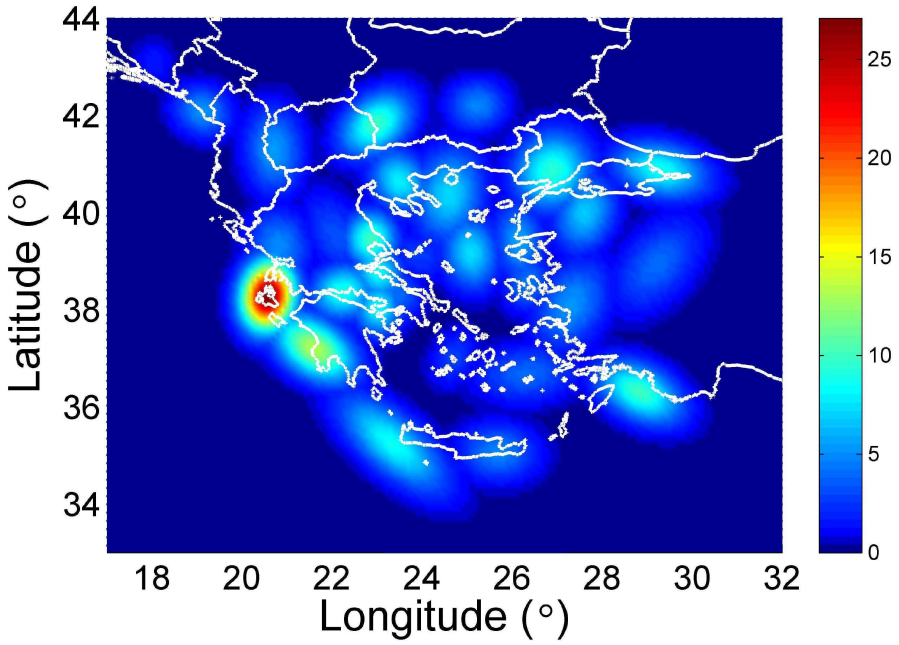

Figure 4.25: Surface of the fit of the 2D Gaussian function to the K = 30 partition of Aegean ruptures, see text for description of the method

For the K = 30 rupture partition the function surface of equation 4.22 is shown in Figure 4.25. The alignment of the Gaussian function for each cluster often aligns closely with the ruptures where the length of the longest rupture is large compared to the distribution of ruptures within the cluster. Where the ruptures are well dispersed within a cluster the Gaussian function appears more circular.

To translate this surface into seismogenic source zones the boundary of each zone is delineated by the location of equal function value between a cluster and its neighbours. This can be done either by inspection or via an algorithm. Here the decision is made to partition the grid of evenly spaced points, as used in the tessellation method ($0.02^\circ \times 0.02^\circ$), and assign each point to a cluster according to which cluster gives a higher value of the smoothed function for the point. For the K = 30 example this is shown in Figure 4.26, with the partition of the rupture catalogue superimposed over it.

The zone model shown in Figure 4.26 clearly bears a significant resemblance to zones delineated using the tessellation method for the K = 30 rupture partition in Figure 4.20d. Whilst the mosaic pattern is not quite so uniform, the general shape of the zones is very

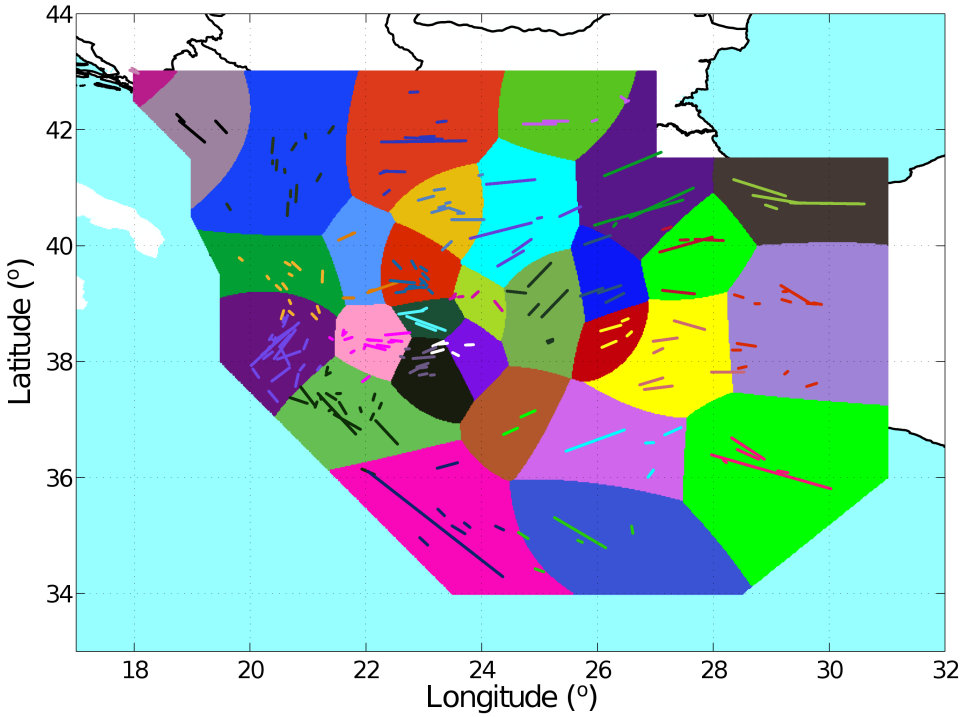

Figure 4.26: Seismic source zones for the K = 30 partition created by splitting neighbouring clusters along lines of equal fit of the 2D Gaussian function. The partitioned ruptures are superimposed on top of the zones

similar. This would suggest that although the tessellation method of delineating uniform zones from the partitions does not explicitly account for the physical effects of faulting, it may serve as a reasonable approximation. Furthermore, the alternative method shown here requires the input of a greater number of parameters. These include the resolution of the smoothing grid and the correlation distance. It is not obvious that the additional uncertainty introduced from these parameters in the alternative method, and the detachment from the observed spatial distribution of seismicity arising from Gaussian smoothing, justifies the additional complexity

4.9.4 Future Prospects for the K-means methodology

The application of K-means cluster analysis to the issue of seismic source modelling is a relatively novel one. The material presented in this chapter represents an initial foundation upon which modifications and extensions can be made to the method. The Aegean region is a useful test region for this particular method. This is because of the quality and time-extent of the historical earthquake catalogue and the extensive literature on the seismotectonics. It is also of great importance because of the diversity of faulting and the mixture of inter- and intra-plate style seismicity. The evolution of seismic source models for this region has been shown in the previous chapter, but the abundance of previous models allows for useful comparison. The existence of a rupture catalogue clearly adds a new dimension to the analysis (both literally and metaphorically), something which is

not always possible elsewhere in the world.

There is considerable disparity across the globe regarding the extent to which seismic sources are well-characterised. This disparity does not necessarily correlate very strongly with the size of the hazard present. For regions such as California, New Zealand and Japan, high seismicity combined with an extensive, well-funded and well-coordinated effort to identify seismogenic faults, have meant that seismic sources, their activity and their dimensions are well-characterised. In other areas such as Central and Southeast Asia, or in low seismicity regions, active faults are not as readily defined. This may be because of a lack of investigation, or because the geography of the region is such that identification of the sources is not always possible. In such regions where the observed seismicity provides the greatest information as to the seismotectonics of a region, the K-means cluster analysis could perhaps prove a more powerful method in the delineation of seismic sources than existing methods. An example of this application can be seen in Burton *et al.* (2008).

In spite of some of the successes of this application, several of the most fundamental issues regarding both cluster analysis and seismic source delineation remain unresolved. Different methods of identifying the optimum number of clusters have been demonstrated here, yet a robust global optimum is not obvious. Equally, for a specified K the global optimum partition has not necessarily been determined either. These are ongoing problems in cluster analysis. Because of the extensive range of applications of cluster analysis in physical and biological sciences, there is continual progress made in addressing them. It is hoped that future developments in the K-means algorithm could also be implemented to help solve the partitioning problem that has been described here.

Some suggestions for improvements to this application of K-means cluster analysis are:

1. The use of stochastic optimisation procedures designed to find the global optimum partition. Some discussion of Genetic K-means procedures was given in Section 4.4.2. They have not been used here largely because the increase in the computational expense is such that they are impractical to run except on the most basic cluster analysis problems. Improvements in processing capabilities, as well as faster heuristics in the algorithms could make them a potential tool in the future. In particular, the Genetic K-medoids algorithm of Sheng and Liu (2006) may prove especially useful, as it combines the searches for the optimum partition as well as the optimum K.
2. Incorporation of uncertainty in hypocentral location and magnitude into the analysis.
3. The use of alternative cluster analysis techniques to elucidate regions of high compact seismicity from dispersed low regional seismicity. In using other clustering techniques to elucidate active sources from low seismicity intra-plate regions, the

K-means algorithm could prove more robust. This is because clusters of high seismicity would be well separated, thus making recognition of the optimum number of clusters easier.

4. Incorporating more information about the dimension and geometry of fault segments into the cluster analysis.

5. Defining other characteristics of faulting in a region into the K-means procedure, e.g. slip rate, stress drop, recurrence intervals of large earthquakes etc.

In parallel with the development of cluster analysis algorithms, ongoing advancement in the field of artificial learning could also have potential for application to the seismic source zonation problem. Eventually this may lead to development of an algorithmic process whereby judgment of better partitions is based on the performance in seismic hazard analysis. Learning procedures such as artificial neural networks may help not only identify the optimum partition, but also the delineation of uniform zones around the partition in accordance with the fit to observed seismicity. The artificial judgement of zonation can be compared with expert judgement as a means of increasing the objectivity of the source zonation, without removing the seismologist's judgement completely.

Chapter 5

Strong Motion Attenuation Relations for Use in the Aegean Region

The translation of earthquake recurrence into strong motion recurrence via empirical attenuation relations is one of the most complex issues in PSHA. Regardless of whether it is the Cornell (1968)-McGuire (1976) method, extreme value analysis or Monte Carlo methods, empirical attenuation relations are needed to describe the decay of ground motion with distance from an earthquake of magnitude M. The selection of candidate attenuation relations is not a trivial procedure, and can often be over-simplified (or at least poorly justified). Several hundred attenuation relations have been published, with new ones constantly emerging. These encompass different regions, countries and tectonic provinces; hence selection of an appropriate model or models requires careful consideration.

The most common ground motion parameter used in seismic hazard analysis is ground acceleration. For mapping applications PGA or zero-period acceleration is preferred. For site-specific applications it may be necessary to consider spectral acceleration. For consistency with previous hazard analyses, acceleration is the preferred parameter here. Some more recent attenuation relations have begun to consider ground velocity and displacement and their respective spectral ordinates. Also, several attenuation relations have been developed for alternative duration-dependent strong motion parameters, which often display a stronger correlation with damage and macroseismic intensity. Attenuation relations with parameters other than acceleration are still not as common as acceleration attenuation relations. It is often the case, therefore, that with fewer relations to choose from, a greater degree of flexibility has to be adopted to allow for their use in the Aegean region.

A set of suggested guidelines (Cotton *et al.*, 2006) has been consulted in the initial stages of selection to identify candidate models. These guidelines include: applicability to the Aegean region, characterisation of soil amplification, characterisation of aleatory vari-

ability, magnitude and distance parameters used, regression methodology, definition of horizontal strong motion (e.g. geometric mean, larger horizontal component, arithmetic mean etc.) and more. An initial suite of models with possible applicability to the Aegean region is presented in this chapter. Justification for rejection of erroneous models is initially given. Those models that are not rejected will be tested against strong motion data for the Aegean, and the goodness of fit analysed using the likelihood method of Scherbaum *et al.* (2004a). For those relations that have not been rejected under both qualitative and quantitative testing, a weighting scheme is suggested to allow for their inclusion into the epistemic uncertainty analysis of the seismic hazard.

Only shallow crustal seismicity is considered at this point. Several intermediate and deep earthquake attenuation relations exist in published literature (Youngs *et al.*, 1997; Atkinson and Boore, 2003; Lin and Lee, 2008), although significantly fewer than for shallow seismicity and, to the author's knowledge, none specific to the Aegean region. Furthermore, Aegean earthquakes are notably absent from the data sets used in the construction of worldwide subduction zone attenuation relations (Youngs *et al.*, 1997; Atkinson and Boore, 2003). This means that they may have a limited applicability to the Aegean region. Indeed, analysis of strong motion from the 2006 Kythira earthquake (M_W 6.7, Focal Depth 60 km) reveals that observed ground accelerations were consistently lower than those predicted from the Youngs *et al.* (1997) relation (Konstantinou *et al.*, 2006).

5.1 Functional form of a Predictive Attenuation Relation

5.1.1 The General Form

Attenuation relations of the sort described in this chapter are generally referred to as predictive relations. They are empirical relationships derived from observed strong motion data. The purpose of these relations is to predict the strong ground motion that would arise from a particular scenario, for example the ground motion on a "rock" site arising from an earthquake of magnitude M at distance R from the site. Consequently the ground motion (defined for the moment as Y) can be represented thus:

$$Y = f(M, R, \theta_i) \pm \sigma \tag{5.1}$$

In this model θ_i refers to an unspecified number of parameters designed to characterise different aspects of ground motion (site response, fault type, wave propagation effects, critical reflections off the Mohorovicic discontinuity etc.). σ is a term to characterise the aleatory variability within the ground motion.

It is well established that peak values of ground motion are lognormally distributed (Kramer, 1996; Douglas, 2003). As such, regressions are usually performed on the logarithm (common or natural) of peak ground motion. It is also recognised that, given

the definition of earthquake magnitude, ground motion should scale proportionally with earthquake magnitude. This commonly results in a linear magnitude scaling:

$$\log Y = c_1 + c_2 M + f(R, \theta) \pm \sigma \tag{5.2}$$

where c_1 and c_2 are constants. There has been an increasing trend in recent attenuation relations to treat the magnitude scaling as a higher order polynomial, or to scale in a more complex manner. This is particularly common in predictive attenuation relations for duration dependent ground motion parameters such as Cumulative Absolute Velocity or Arias Intensity.

5.1.2 Attenuation with Distance

As seismic waves radiate away from the earthquakes source they undergo attenuation in two forms. The first is geometric spreading, where waves lose energy as they radiate over an increasing area. The second is intrinsic damping, where energy is absorbed by internal friction within the medium of travel. Surface wave amplitudes tend to decay according to $R^{-1/2}$, hence the geometric spreading term is usually included as $\log R$. The intrinsic damping term is often seen as a simple decay in $\log Y$ with R, and is often found not to be a significant regression parameter, especially over shorter distances. Generally attenuation scaling with distance is usually seen as:

$$\log Y = c_1 + c_2 M + c_3 R + c_4 \log(R) + f(\theta) \pm \sigma \tag{5.3}$$

The definition of distance R is an area of considerable complexity, hence the generic term R here simply reflects increasing distance from the source. R is usually defined as the vectorial mean of source to site distance on the Earth's surface and the source depth. Depending on the data used in the regression, c_3 may not be significant. The scaling of ground motion with distance is often even more complex than this simple decay relation. As the area of fault rupture scales with magnitude, it is often observed that waves arriving at a site will originate from a range of distances, which increases with increasing rupture length [magnitude]. It is not uncommon, especially in newer relations, to define a non-linear magnitude-distance scaling term, often scaling linearly with magnitude. Consequently, the attenuation may take the form of:

$$\log Y = c_1 + c_2 M + c_3 R + (c_4 + c_4 M) \log(R) + f(\theta) \pm \sigma \tag{5.4}$$

Or in more complex cases:

$$\log Y = c_1 + c_2 M + c_4 R + g(M, R) + f(\theta) \pm \sigma \tag{5.5}$$

Where g(M,R) is a non-linear function of magnitude and distance.

5.1.3 Soil conditions and style-of-faulting

The remaining term θ describes parameters to account for soil amplification and the impact of fault type on the ground motion. It is common, even in more recent attenuation relations for $\log(Y)$ to scale with fault type or site class in a linear fashion.

$$f(\theta) = c_6 S_A + c_7 S_S + c_8 F_N + c_9 F_R \tag{5.6}$$

This example defines three types of site condition, usually given in terms of NEHRP Site Classification (FEMA, 2003). $S_A = 1$ refers to stiff-soil or alluvium (NEHRP Class C); $S_S = 1$ to soft soils (NEHRP Class D) and both equal to zero for rock (NEHRP Class B). Depending on the metadata available for the strong motion records, some attenuation relations may prefer to model site amplification directly as a function of 30-m average shear wave velocity (V_{S30}).

There is considerable variation in the style-of-faulting parameters used in different relations. In equation 5.6, $F_N = 1$ for a normal fault, 0 otherwise; $F_R = 1$ for a reverse fault, 0 otherwise. Generally, it is observed that earthquakes that display reverse or reverse-oblique slip tend to produce higher ground motions than strike-slip events, whilst normal faulting earthquakes produce lower ground motions. Style-of-faulting is determined using the earthquake rake. The classification of types of faulting using rake often varies for different relations (Bommer *et al.*, 2003). It may be the case that the style-of-faulting terms consider only two faulting categories, sometimes combining reverse and strike slip into one parameter, or combining normal and strike slip into one parameter. Alternatively, the authors may choose to separate the style-of-faulting into four or more categories. The inconsistency in the definition of style-of-faulting and the parameters used makes comparison of attenuations within an epistemic uncertainty analysis more complicated.

There has, in recent attenuation relations, emerged a growing trend for incorporating non-linear soil amplification into the attenuation model (Choi and Stewart, 2005). These models tend to require calculation of ground motion on a reference rock site, and then amplify according to the strength of the ground motion on rock. Similar increases in complexity can be seen for style of faulting terms.

5.1.4 Aleatory Variability

The σ term in equation 5.1 is a parameter to account for the scatter of the residuals of $\log(Y)$ around the median strong motion attenuation curve. This represents the natural variability of ground motion, otherwise referred to as aleatory variability. Earlier attenuation relations using two-step regression (Joyner and Boore, 1981) tended to treat σ as a single parameter. The emergence and widespread adoption of a random effects regression procedure (Brillinger and Preisler, 1984; Abrahamson and Youngs, 1992) has

allowed for distinction between inter- (τ) and intra-event (ε) variability. These terms are then combined to give a total aleatory variability term $\sigma = \sqrt{\tau^2 + \varepsilon^2}$. The importance of aleatory variability in seismic hazard analyses is well recognised (Bommer and Abrahamson, 2006). There are many developments in the understanding of aleatory variability that are still the subject of debate. Some of these will be covered later in this chapter.

The general functional form of a ground motion attenuation relation has been described. A full treatise of the complex issues in the development of predictive ground motion relations is beyond the scope of this thesis, though the reader is referred to Douglas (2003) for such a treatment. Several major issues will be addressed in this chapter when considering particular relations. There are still several topics that require some further consideration.

5.1.5 Selection of ground motion data

The transformation of strong motion data, from the form in which it is processed by an accelerograph to the spectrum that is used in the regression, is complicated. It can strongly influence the form of the predictive equation. Ground motions below a threshold level (typically in the 0 - 20 cm s^{-2} range) are excluded due to the variability in trigger threshold of the strong motion instrument. This does, however, create a magnitude-distance correlation in the data set, as fewer records are considered from smaller earthquakes at distance. Records are also filtered to exclude frequencies typically lower than 0.1 - 0.2 Hz and higher than 30 Hz. The type of filter used does vary considerably, which can impact on the spectral recording at the extreme ends of the frequency range.

Strong motion records typically come from three component broadband instruments. This provides two components of horizontal ground motion. Exactly how horizontal ground motion is defined in the data set does, again, vary from relation to relation. Common approaches are to use either the arithmetic or geometric mean of the two spectral acceleration data sets, to use the larger component, a random component, or sometimes the vectorial mean. The choice of method will strongly influence the regression equation, especially with respect to the σ value. This also adds further error into the analysis of epistemic uncertainty when comparing several relations.

5.1.6 Source and Distance Metrics

The characterisation of the source and distance are yet further parameters that vary from attenuation model to attenuation model. Most attenuation relations developed in the last 10 to 15 years use moment-magnitude (M_W) as the preferred parameter. Prior to this the standard magnitude parameter was typically surface-wave magnitude M_S or Richter local magnitude (M_L). The latter magnitude scale is still used occasionally in attenuation relations for regions with a highly localised data set. In this work the earthquake catalogue has been homogenised into a standard moment magnitude scale. Consequently, it

is M_W that is simulated in the synthetic catalogues, which means application of M_S or M_L dependent relations requires conversion of the earthquakes to the appropriate magnitude scale. Theoretically, the uncertainty in the conversion between magnitude scales can be incorporated into the aleatory variability term. The M_W to M_S scaling relation used in chapter 2, however, would suggest that this may not be necessary for the magnitude range being considered here ($M_W > 5.0$).

The distance metrics are considerably more complicated and several different metrics are in widespread usage. The most common types of distance metrics observed are epicentral distance (R_{EPI}), hypocentral distance (R_{HYP}), Joyner-Boore distance (R_{JB}) and rupture distance (R_{RUP}). The first two refer to the distance from the site to the epicentre (across the ground surface) and the hypocentre respectively. Joyner-Boore distance is defined as the distance from the site to the nearest surface projection of the rupture plane. Rupture distance is defined as the distance from the site to the nearest point of the seismogenic rupture. These metrics are illustrated in Figure 5.1.

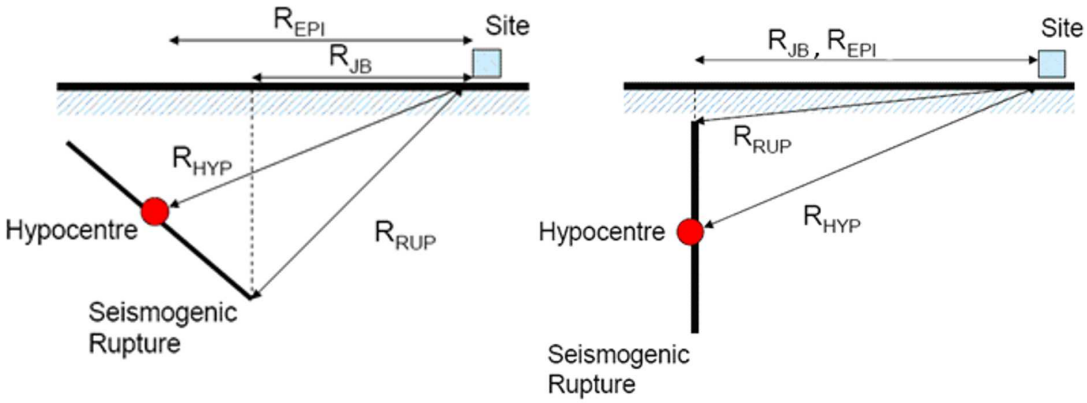

Figure 5.1: 2D schematics demonstrating the difference in definitions of source to site distance for a) Non-vertical dipping fault and b) a strike-slip fault

Figure 5.1 clearly illustrates that these metrics are not interchangeable, nor can conversion be made via a simple geometric transformation unless the dimensions of the fault are well known. In the Aegean region many seismogenic sources, both offshore and onshore are poorly mapped (Danciu and Tselentis, 2007). This has been discussed at length in previous chapters. Attenuation relations derived from Greek earthquakes alone tend to use epicentral or hypocentral distance as the preferred metric. Elsewhere, particularly in California, Europe and Japan, Joyner-Boore distance is the preferred metric for shallow crustal events, and rupture distance for deeper subduction events.

In the Monte Carlo simulations presented in this work, hypocentres are determined from a stochastic point process. Therefore, epicentral and hypocentral distances (R_{EPI} and R_{HYP} respectively) are the simplest and most computationally efficient to determine within the Monte Carlo simulation. Empirical equations are available for conversion between different metrics. These are presented in section 5.6.

5.2 Candidate Attenuation Relations: Peak Ground Acceleration

A full list of published attenuation relations up to the year 2001 (Douglas, 2003) has been consulted for initial identification of possible candidate relations. For 2001 to the present, attenuation relations have been identified by literature search. Most published relations were immediately rejected using the criteria of Cotton *et al.* (2006). This was done on the basis of the region (e.g. subduction zones, intraplate interiors etc), magnitude scale (e.g. M_{JMA}, M_L), inappropriate functional form and age or limitations of records. Ten models for PGA, five with spectral ordinates, will be considered in further detail.

Each attenuation model will have a specific magnitude and distance range over which they are applicable. They may also have different magnitude and distance metrics, soil and fault characterisation and aleatory uncertainty term. For spectral attenuation relations, each relation may have a different applicable spectral range. These ranges, and other relevant information is summarised in Table 5.1. The magnitude and distance ranges given in Table 5.1 are those suggested by the authors themselves. The σ column refers to the characterisation of aleatory variability, which may be either absent (none), a single parameter (single) or an intra- and inter- event term (two).

(a) PGA attenuation [1]Error in site classification for strong motion records used in analysis

Attenuation Model	Region	No. of Records	No. of Earthquakes	Regression	Horizontal Component	M	M_{LOW}	M_{HIGH}	R	R_{LOW} (km)	R_{HIGH} (km)	Site Class.	Fault	σ
Joyner and Boore (1981)	California	182	23	2-Stage	Larger	M_W	5	7.7	R_{JB}	0.5	350	None	None	Single
Makropoulos and Burton (1985b)	Unknown	N/A	N/A	N/A	N/A	M_S	U	U	R_{HYP}	U	U	None	None	None
Theodulidis and Papazachos (1992)	Greece	121	40	2-stage	Independent	M_S	4.5	7	R_{EPI}	1	130	B,C/D[1]	None	Single
Ambraseys (1995)	Europe & Middle East	1260	619	2-Stage	Larger	M_S	4	7.3	R_{JB}	1	200	None	None	Single
Ambraseys *et al.* (1996)	Europe & Middle East	1260	619	2-stage	Larger	M_S	4	7.3	R_{JB}	1	200	B,C,D	None	Single
Margaris *et al.* (2001)	Greece	474	142	2-stage	Unknown	M_W	4.5	7	R_{EPI}	5	120	B,C,D	None	Single
Skarlatoudis *et al.* (2003)	Greece	1000	225	2-stage	Unknown	M_W	4.5	7	R_{EPI}	5	160	B,C,D	2 (N,SS/T)	Single
Ambraseys *et al.* (2005a)	Europe & Middle East	595	135	Random Effects	Larger	M_W	5	7.6	R_{JB}	0	100	B,C,D	4 (SS, N, R, O)	Two - M dependent
Danciu and Tselentis (2007)	Greece	335	151	Random Effects	Geometric Mean	M_W	4.5	6.9	R_{EPI}	0	136	B,C,D	2 (N,SS/R)	Two
Bommer *et al.* (2007)	Europe & Middle East	997	289	Random Effects	Larger	M_W	3	7.6	R_{JB}	0	100	B,C,D	3 (N,SS,R)	Two - M dependent
Boore and Atkinson (2007)	Worldwide	3552	175	Random Effects	Orientation-Independent Geometric Mean	M_W	5	8	R_{JB}	0	200	V_{s30}	4 (U, N, SS, R)	Two

(b) Spectral attenuation (T_N refers to the number of spectral ordinates modelled)

Attenuation Law	T_N (s)	T_{LOWER} (s)	T_{UPPER} (s)	$FILTER_{LOWER}$ (s)	$FILTER_{UPPER}$ (s)	FILTER TYPE
Ambraseys *et al.* (1996)	46	0.1	2	0.05	5	Bandpass
Ambraseys *et al.* (2005a)	61	0.05	2.5	0.04	2.5	Butterworth
Danciu and Tselentis (2007)	31	0.1	4	0.04	4	Bandpass
Bommer *et al.* (2007)	11	0	0.5	/	4	Low pass
Boore and Atkinson (2007)	22	0.01	10	Variable	Variable	Variable, mostly Butterworth

Table 5.1: Properties of the attenuation models considered for Greece

5.2.1 Joyner and Boore (1981)

This relation has been one of the most widely used PGA attenuation functions in seismology. Most of these strong motion records are from California, with just two from Alaska. Consequently most of the earthquakes are strike-slip or normal faulting events. The resultant equation is given as follows:

$$\log(Y) = -1.02 + 0.249M_W - \log r - 0.00255r + 0.26P \tag{5.7}$$

Where P = 0 for the 50th percentile, and 1 for the 84^{th} percentile ground motion and $r = \sqrt{R_{JB}^2 + 7.3^2}$. The earthquakes used in the data set are all shallow (depth $<$ 20 km). Y is the PGA expressed as a fraction of g. This relation can be seen plotted in Figure 5.2.

This was one of the first attenuation relations to include an explicit aleatory uncertainty value. The lack of significant difference between rock and soil sites is surprising, and may imply more about the distribution of strong motion recordings on soft soil sites than it does about the nature of strong motion at such sites.

It has been suggested by Burton *et al.* (2001) that this could be applied to the Aegean region, despite no Aegean earthquakes contributing to the data set. Much of this assumption derives from the tectonic similarities between California and the dextral strike-slip regime of the North Anatolian Fault. To what extent this could be applied across the entire Aegean, however, is still a matter of debate. It may be the case that the relation is a valid approximation for normal and strike-slip earthquakes away from the Eurasian plate margin, but should not be applied to thrust or subduction earthquakes along the Hellenic arc.

5.2.2 Makropoulos and Burton (1985b)

This attenuation model is distinct from all the others considered here, as it is not derived directly from a set of strong motion data, but from an average of eight existing relations; all that were known to be specific to Greece in the early 1980s. Most of these relations are derived for hard rock sites exclusively, or for sites with very shallow soils. This means that they can only be used for general and mapping purposes, rather than site-specific. The Makropoulos and Burton (1985b) relation has no single uncertainty value and no site condition attached, although uncertainties on the magnitude and distance terms are given. The relation has the functional form:

$$Y = 2164e^{0.7(\pm 0.03)M_s}\left(R_{HYP} + 20\right)^{-1.80(\pm 0.02)} \tag{5.8}$$

where Y is the PGA given in cm s^{-2}. It is unclear over what range of depths this relation is applicable. This relation is plotted in Figure 5.2. Presented in the paper are comparisons of the "average" predicted PGA with eight observed PGA recordings from

earthquakes in the Aegean. There is reasonable agreement between the predicted and observed accelerations although the magnitudes and distances of the tested earthquakes span a relatively small range ($4.3 \leq M_S \leq 5.9$; $20 \leq R_{HYP}\,(km) \leq 56$). No indication is given regarding the ranges of magnitudes and distances over which the relation should be applied.

It is very difficult to determine whether this relation should still be used in seismic hazard analysis in the Aegean region. It has been used by Makropoulos and Burton (1985b) and Burton *et al.* (2003), and comparisons of predicted and observed ground motion are surprisingly consistent. The difficulty in characterising a total aleatory variability parameter may cause problems in the application. Similarly the absence of a site coefficient limits its use. When using the selection criteria of Cotton *et al.* (2006) this relation is a candidate for rejection on many grounds. Most of these relate to the absence of information regarding the data used in its derivation. However, it should also be noted that this relation is derived in a very different way to the "ideal" attenuation relation that would pass all of the selection criteria. Consideration may be given for the purposes of seismic hazard mapping, but not for use in site-specific studies.

5.2.3 Theodulidis and Papazachos (1992)

This attenuation relation is one of the first derived from Greek earthquakes that includes PGV and PGD. In addition, the authors suggest empirical relations between peak horizontal ground motion and Modified Mercalli Intensity (I_{MM}). The acceleration relation takes the following form:

$$\ln(Y) = 3.88 + 1.12M_S - 1.65\ln(R_{EPI} + 15) + 0.41S + 0.71P \tag{5.9}$$

where, Y is acceleration in cm s^{-2} and S equal to 0 at alluvium sites and 1 at rock sites. This produces the surprising result that ground acceleration is greater on rock sites than on soil sites, though velocity and displacement are lower on rock sites than on alluvium. The authors explain this as a result of soils concentrating energy in lower frequencies, thus favouring greater displacements, and rock concentrating energy in the higher frequencies, favouring greater accelerations. However, reinvestigation of the strong motion sites (described in Burton *et al.* (2003)) found that sites previously characterised as "rock" had been found to be perturbed by a thin layer of weathered material that amplified the acceleration. It has been subsequently recommended that an "intermediate" term of S = 0.5 ('stiff' soil) be used in place of the site parameter, thus resulting in:

$$\ln(Y) = 4.09 + 1.12M_S - 1.65\ln(R_{EPI} + 15) \tag{5.10}$$

Note, however, that readjustment of the P value for this compensation is not given; hence this adaptation can only be used for 50th percentile acceleration. This figure is plotted alongside the Joyner and Boore (1981) and Makropoulos and Burton (1985b) PGA atten-

uation relations in Figure 5.2.

The Theodulidis and Papazachos (1992) relation generally predicts higher ground motions than other attenuation relations. In the extreme near-field this may exceed 1g even for the 50th percentile relation (Figure 5.2) of an M_W 7 earthquake. This may be partly due to the error in site classification. It is, however, compatible with the tectonics of the Aegean region. The data set used for this relation is made explicit, although there does appear to be a weak correlation between distance and magnitude, which is not addressed in the paper.

This relation would pass all of the selection criteria, were it not for two problems. The first is the conversion between the magnitude scales and the second is the error arising from incorrect site classification. Use of the adjusted formula gives cause for concern as it is not made clear how this adjustment affects the aleatory variability term σ.

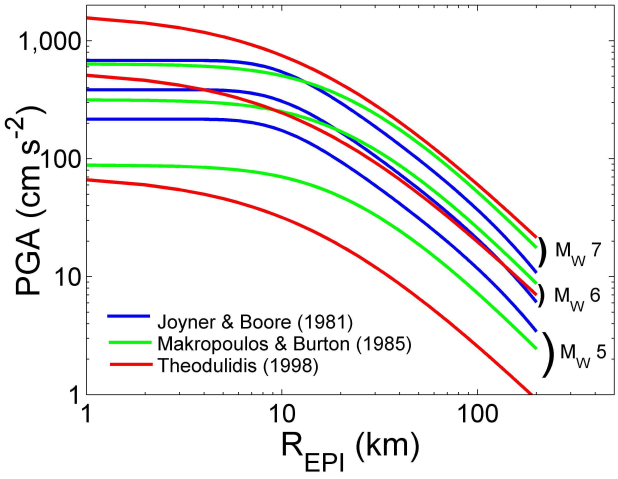

Figure 5.2: Comparison of the Joyner and Boore (1981) (blue),Makropoulos and Burton (1985b) (green) and adjusted Theodulidis (1998) (red) PGA attenuation models (50^{th} percentile). Depth is assumed to be 10 km

5.2.4 Ambraseys (1995)

This attenuation model for PGA has, until recently, been amongst the most extensively used for its applicable region (Europe and the Middle East). Approximately 15 % of the data are from Greece, which represents a significant proportion of the data set. Several attenuation relations are produced, which include horizontal acceleration and vertical acceleration with or without depth control. The preferred equations are:

$$\log(Y_H) = -1.429 + 0.245M_s - 0.0010r - 0.786\log(r) + 0.241P \tag{5.11}$$

This equation refers to the horizontal acceleration, without depth control. Here, Y_H is the horizontal PGA (in g), P = 0 for the 50^{th} percentile and 1 for the 84^{th} percentile, and

$r = \sqrt{R_{JB}^2 + h_0^2}$ where $h_0 = 2.7$ km. With depth control, h_0 refers to the focal depth of the earthquake. In this case equation 5.11 becomes:

$$\log(Y_H) = -1.060 + 0.245M_s - 0.00045r - 1.016\log(r) + 0.25P \tag{5.12}$$

The formula for vertical PGA is:

$$\log(Y_v) = -1.72 + 0.243M_s - 0.00174r - 0.750\log(r) + 0.24P \tag{5.13}$$

This formula is derived without depth control, consequently $h_0 = 1.9$. With depth control this becomes:

$$\log(Y_v) = -1.33 + 0.248M_s - 0.00110r - \log(r) + 0.25P \tag{5.14}$$

Ambraseys (1995) also compares these with equations derived using one-stage regression, the differences are small across much of the range and the two stage regression is preferred.

One particular shortcoming common to these formulae is the lack of a site parameter. This is justified by referring to Ambraseys and Bommer (1992), who investigated the site conditions and found that the differences between the two classes ('rock' and 'soil') were insignificant for two stage regression given the data base used. However, further regressions were undertaken using both one- and two-stage regressions with local shear-wave velocity V_S used as a site parameter. For a subset of 268 strong motion records from 132 earthquakes, soil profile conditions are known and the average shear-wave velocity for 30 m depth of soil (V_{S30}) is used as the site parameter in the regression. Without depth control, this produces, for one-stage regression:

$$\log(PGA_H) = -1.31 + 0.273M_S - 0.781\log(r) - 0.12\log(V_{s30}) + 0.238P \tag{5.15}$$

For two-stage regression this becomes:

$$\log(PGA_H) = -1.05 + 0.245M_S - 0.001r - 0.786\log(r) - 0.15\log(V_{s30}) + 0.23P \tag{5.16}$$

Without knowledge of the local site conditions it can be difficult to use the attenuation relation in 5.16 to predict PGA. This is also compounded by the fact that the strong motion records taken on known soil profiles span a much smaller range of distances and magnitudes than those without dependence on V_{s30}.

The extensive data set used by Ambraseys allows for a large magnitude and distance range. Generally, these attenuation relations are good candidates for application to the Aegean region. The absence of site parameter is also unrealistic, though new methods of calculating V_{S30} (Wald and Allen, 2007) make it possible to use the latter relations in both site-specific hazard analysis and hazard mapping. Comparison between this relation

(equation 5.12) and the later "Ambraseys" relations is shown in Figure 5.3.

5.2.5 Ambraseys *et al.* (1996)

This study is an extension of the Ambraseys (1995) study designed to develop spectral attenuation relations for Europe and the Middle East. More consideration to this will be given when discussing spectral attenuation relations. In order to be consistent with the spectral ordinate regressions, however, an additional PGA regression was undertaken. Unlike the Ambraseys (1995) relation, here site-conditions are known for 416 strong motion records. These are broken down on the basis of soil category into NEHRP sites classes B, C, and D. Using the same magnitude and distance parameters as in Ambraseys (1995) the new regression becomes:

$$\log(Y_H) = -1.48 + 0.266M_S - 0.922\log(r) + 0.117S_A + 0.124S_s + 0.25P \tag{5.17}$$

Where S_A is 1 for class C, 0 otherwise, and S_S is 1 class D, 0 otherwise. This relation can be seen alongside the Ambraseys (1995) relation in Figure 5.3. The same data set produces the following relation for vertical acceleration:

$$\log(Y_V) = -1.74 + 0.273M_S - 0.954\log(r) + 0.076S_A + 0.058S_S + 0.26P \tag{5.18}$$

5.2.6 Margaris *et al.* (2001)

This study of attenuation for Greek earthquakes is the only one that was not published in peer reviewed literature. Empirical regression equations are presented for peak ground acceleration, velocity and displacement. The earthquakes display mostly normal faulting; hence no fault parameter is included in the relation. Threshold criteria are that $M_W \geq 4.5$, $PGA \geq 5\ cm\ s^{-2}$ for at least one record from each earthquake. The authors note some correlation between magnitudes and distance in the data set.

The site effects are taken into account by virtue of a parameter S, which corresponds to 0 for NEHRP class B soil, 1 for class C soil and 2 for class D soil. For each type of ground motion, two equations are presented, the first uses average focal depth ($h_0 = 6$ km), the second uses average "effective" depth ($h_0 = 7$ km):

$$\ln(Y) = 4.16 + 0.69M_W - 1.24\ln(R_{EPI} + 6) + 0.12S \pm 0.70P \tag{5.19}$$

$$\ln(Y) = 3.52 + 0.70M_W - 1.14\ln\left(\sqrt{\left(R_{EPI}^2 + 7^2\right)}\right) + 0.12S \pm 0.70P \tag{5.20}$$

The validity range of these relations are comparatively small when viewed alongside the Ambraseys (1995) and Ambraseys *et al.* (1996) relations. Furthermore, as the data set was comprised of mostly normal faulting events, with some strike-slip events, it would have

limited use along the Hellenic arc and Adriatic coast, where thrust faulting is dominant. Although this relation does pass most of the Cotton *et al.* (2006) criteria for accepting an attenuation relation, the narrow range and the dependence on normal faulting limit its use across the entire Aegean. This relation is compared alongside other Greek-specific relations (Skarlatoudis *et al.*, 2003; Danciu and Tselentis, 2007) in Figure 5.4.

5.2.7 Skarlatoudis *et al.* (2003)

This study is largely a development of the Margaris *et al.* (2001) regression equations with an expanded data set. As with the Margaris *et al.* (2001) data set, these are mostly normal and strike-slip events; however, 67 thrust events are included in the data set. The same strong motion record selection criteria as listed in Margaris *et al.* (2001) are used, though once again it is unclear whether the greater of the two horizontal components or the mean of the components is used as the regression variable.

The key difference between this and the Margaris *et al.* (2001) relation is the inclusion of a term to characterise the fault type in the regression. Here F = 0 for normal faulting events and F = 1 for strike-slip and thrust events. Two equations for each ground motion variable are presented, the first uses focal depth as a free parameter, the second uses average focal depth (h_0 = 6 km) as a constraint.

$$\log_{10}(Y) = 0.86 + 0.45M_W - 1.27\log_{10}\left(\sqrt{\left(R_{EPI}^2 + h^2\right)}\right) + \cdots$$
$$0.10F + 0.06S + 0.286P \tag{5.21}$$

$$\log_{10}(Y) = 1.07 + 0.45M_W - 1.35\log_{10}\left(R_{EPI}^2 + 6\right) + \cdots$$
$$\cdots + 0.09F + 0.06S + 0.286P \tag{5.22}$$

The S term is equal to 0 for NEHRP class B, 1 for class C and 2 for class D. Though the coefficients differ from the Margaris *et al.* (2001) equation, the magnitude validity range is the same. The distance range has increased slightly. The inclusion of a term to characterise thrust faulting and strike-slip faulting would suggest that this relation could be used across much of the Aegean. In particular it may be useful for the Hellenic Arc and Adriatic coast. This relation is compared with the Margaris *et al.* (2001) equation in Figure 5.4.

5.2.8 Ambraseys *et al.* (2005a)

A PGA relation is included in the spectral attenuation relations of Ambraseys *et al.* (2005a, 2005b). PGA attenuation relations can be found for horizontal acceleration Ambraseys *et al.* (2005a) and vertical acceleration Ambraseys *et al.* (2005b).

$$\log\left(Y_h\right) = 2.522 - 0.142M_W + \left(-3.184 + 0.314M_W\right)\log\left(\sqrt{R_{jb}^2 + 7.6^2}\right) + \cdots$$
$$\cdots + 0.134S_S + 0.050S_A - 0.084F_N + 0.062F_T - 0.044F_0 + \sigma P \tag{5.23}$$

Here S_S refers to stiff soil (1 if true, 0 otherwise), S_A soft soil, F_N normal faulting earthquakes, F_T thrust faulting earthquakes, F_O oblique faulting earthquakes. Uncertainty is given in terms of both inter- and intra-event variability. Both of these terms are magnitude dependent, with σ defined as $\sigma = \left(\left(0.665 - 0.065M_W\right)^2 + \left(0.222 - 0.022M_W\right)^2\right)^{1/2}$.

The vertical acceleration is defined as:

$$\log\left(a_v\right) = 0.835 + 0.083M_W + \left(-2.489 + 0.206M_W\right)\log\left(\sqrt{r_{jb}^2 + 5.6^2}\right) + \cdots$$
$$\cdots + 0.078S_S + 0.046S_A - 0.126F_N + 0.005F_T - 0.082F_o + 0.280P \tag{5.24}$$

However, it should be noted that the data set contains very few records from strike-slip, thrust and oblique earthquakes at distances of less than 5 km. Similarly there are no records from normal earthquakes greater than M_W 7. This relation is plotted alongside the Ambraseys (1995) and Ambraseys *et al.* (1996) relations in Figure 5.3.

These relations are markedly different from their European and Mid-Eastern predecessors found in Ambraseys (1995) and Ambraseys *et al.* (1996). Clearly the characterisation of fault type and site response is more sophisticated, though this does not necessarily reflect an improvement of the regression. This much is evident from the aleatory variability, which is similar to that found in Ambraseys (1995) for moderate sized earthquakes. The dependence of the variability on magnitude may cause problems in the practical application of this algorithm. When this relation is applied to a simple PSHA the hazard curve remains largely insensitive to the M_{MAX} value (Musson, 2009). This is because of the increasing aleatory variability with smaller magnitudes. It remains to be seen how well these zero-period accelerations perform in practical application. Their derivation from European earthquakes, in which Aegean region earthquakes are well represented (22 % from Turkey and 19 % from Greece), means that they are considered to be applicable for the Aegean region. Equally, the use of M_W as the magnitude reduces error from conversion. Some error may arise when considering Joyner-Boore distance for earthquakes with magnitudes greater than M_W 6. The distance range is limited, however, and this could

prove to be problematic.

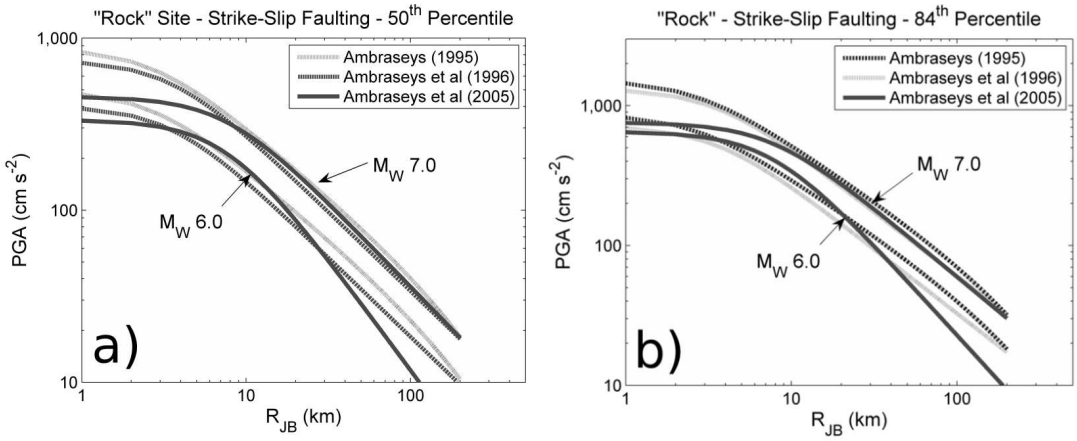

Figure 5.3: Comparison of the Ambraseys (1995), Ambraseys *et al.* (1996) and Ambraseys *et al.* (2005a) attenuation relations for a rock site, with an assumed strike-slip fault. a) 50^{th} percentile and b) 84^{th} percentile.

The Ambraseys *et al.* (2005a) relation predicts lower accelerations in the near field ($R_{JB} < 10\ km$) than its predecessors. This is due to the increased number of records from near-field events in the strong motion data set. For large earthquakes there is reasonable agreement between all three relations for distances greater than 10 km. The major contrast between the Ambraseys *et al.* (2005a) relation and its predecessors is the greater attenuation for smaller earthquakes and the increase in higher percentile accelerations. For the 50^{th} percentile the difference in acceleration between an M_W 6.0 and M_W 7.0 earthquakes at distances of less than 10 km is of the order of 150 to 200 cm s^{-2}. For the 84^{th} percentile this difference has dropped to 75 - 125 cm s^{-2}. This is a clear manifestation of the magnitude dependent variability observed in the Ambraseys *et al.* (2005a) relation.

5.2.9 Danciu and Tselentis (2007)

This attenuation relation is one of the first that considers a large variety of engineering related strong motion parameters for Greece. In addition to PGA and PGV and their respective spectral ordinates, empirical attenuation relations have been determined for Arias Intensity (I_a), root mean square acceleration (a_{rms}), Characteristic Intensity (I_C), Fajfar's Intensity (I_f), Housner Spectrum Intensity (SI), Cumulative Absolute Velocity (CAV) and CAV integrated at 5 cm s^{-2} lower threshold (CAV_5).

Site response is characterised by the parameter S, where S = 0 for rock ($V_{s30} > 750\ m/s$), S = 1 for stiff soil ($360 \leq V_{s30}\ (m/s) \leq 750$) and S = 2 for soft soil ($200 \leq V_{s30}\ (m/s) \leq 360$). Faulting is characterised by parameter F, where F = 0 for normal faulting and F = 1 for strike-slip and thrust faulting. The equation for horizontal acceleration is:

$$\log_{10}(Y) = 0.883 + 0.458 M_W - 1.278 \log_{10}\left(\sqrt{R_{EPI}^2 + 11.515^2}\right) + 0.038S + 0.116F + 0.291P$$

(5.25)

Compared to the relations of Ambraseys (1995), Ambraseys *et al.* (1996) and Ambraseys *et al.* (2005a) the validity range is quite limited (Table 5.1), albeit the relation is derived using exclusively Greek data.

Of all the relations presented here, the Danciu and Tselentis one may appear to be the most suitable for the purposes of Monte Carlo seismic hazard analysis in the Aegean region. The characterisation of earthquakes as a point source is convenient given the inconsistency in the identification of active faults in Greece. Furthermore, the basic Monte Carlo procedure tends to assume a point source. Similarly, moment magnitude is the preferred magnitude scale, which avoids error converting to M_S, the preferred scale for the earlier relations. Clearly the attenuation relation is applicable to Greece, and although most of the selected earthquakes come from normal faulting regimes, several earthquakes in the data set display thrust characteristics associated with the Hellenic subduction. There are, however, no events from Albania, Macedonia, Bulgaria or Western Turkey, and few events from the orogenic collision zone of the Dinaric Alps or from the western North Anatolian fault. As these regions are not represented in the data set, caution should be taken in applying this equation outside of Greece.

Comparisons of the three Greek specific relations Margaris *et al.* (2001); Skarlatoudis *et al.* (2003); Danciu and Tselentis (2007) are shown in Figure 5.4. It is immediately noticeable that all three relations are similar when R_{EPI} is greater than 10 km. The increased number of near-field recordings in the Danciu and Tselentis (2007) data set has placed an obvious constraint on the accelerations observed at epicentral distances of less than 10 km. The Margaris *et al.* (2001) relation also appears to predict lower accelerations than the other relations.

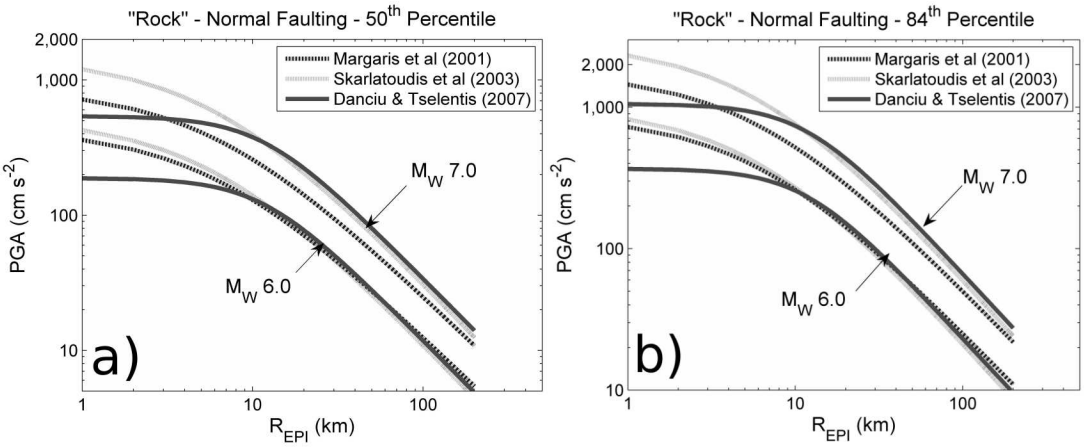

Figure 5.4: Comparison of the Margaris *et al.* (2001), Skarlatoudis *et al.* (2003) and Danciu and Tselentis (2007) attenuation relations for a rock site and normal fault. a) 50^{th} percentile and b) 84^{th} percentile

5.2.10 Bommer *et al.* (2007)

This attenuation relation is derived using mostly the same dataset as that of Ambraseys *et al.* (2005a). It is constructed for the purposes of extending application to low-moderate seismicity regions. It has been recognised that extrapolations of attenuation relations derived from data sets of large earthquakes ($M_W > 5$), overestimate ground motions from small earthquakes. This appears to be the case when the attenuation relation is implemented for earthquakes at the extreme end of the distance range, as well as those earthquakes beyond the valid distance range. Consequently, regression over a larger data set is undertaken. The strong motion data set in Bommer *et al.* (2007) consists of 997 accelerograph recordings from 289 events spanning the magnitude range $3.0 \leq M_W \leq 7.6$, and the distance range $0 \leq R_{JB} \leq 100$ km. This relation is derived for spectral ordinates. The zero period acceleration attenuation relation is presented here:

$$\begin{aligned} \log_{10}(PSA_{T=0}) &= 0.0031 + 1.0848M_W - 0.0835M_W^2 + \cdots \\ &\quad + (-2.4423 + 0.2081M_W)\log_{10}\sqrt{R_{JB}^2 + 8.0282^2} + \cdots \\ &\quad + 0.0781S_S + 0.0208S_A - 0.0292F_N + 0.0963F_R \end{aligned} \tag{5.26}$$

S_S and S_A are coefficients describing the soft soil and stiff soil conditions respectively, and F_N and F_R are coefficients describing the fault type (normal and reverse respectively). This relation is plotted in Figure 5.5.

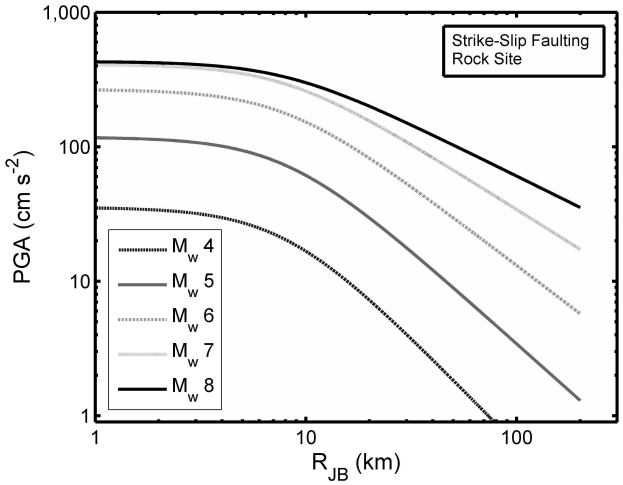

Figure 5.5: Median PGA attenuation over the extended magnitude range using the Bommer *et al.* (2007) relation for a rock site, with strike slip faulting assumed

Aleatory variability in these relations is considered to be heteroscedastic (magnitude-dependent), and for the zero-period relation is defined as $\sigma_T = \sqrt{\sigma_1^2 + \sigma_2^2}$, where $\sigma_1 = 0.599 - 0.058M_W$ (intra-event variability) and $\sigma_2 = 0.323 - 0.031M_W$ (inter-event variability). This magnitude dependent variability is consistent with the aleatory vari-

ability given in the Ambraseys *et al.* (2005a) relation. This will reduce the impact of larger earthquakes on the hazard curve. Since variability increases with decreasing magnitude it is likely that smaller earthquakes with a higher variability will come to dominate the hazard at a site. Generally, the trend of decreasing variability with increasing magnitude is a difficult one to determine and is not always statistically significant. Furthermore, as noted by Musson and Sargeant (2007) the increasing variability at lower magnitudes may also arise from the bias due to the smaller number of records available from smaller earthquakes. A comparison of the 50^{th} and 84^{th} percentiles of ground motion is shown in Figure 5.6, alongside a comparison of ground motion on different soil types. Of note is the saturation of PGA at high magnitudes in the extreme near-field. This feature results from the non-linear magnitude scaling term and the magnitude-dependent distance scaling term shown in equation 5.26.

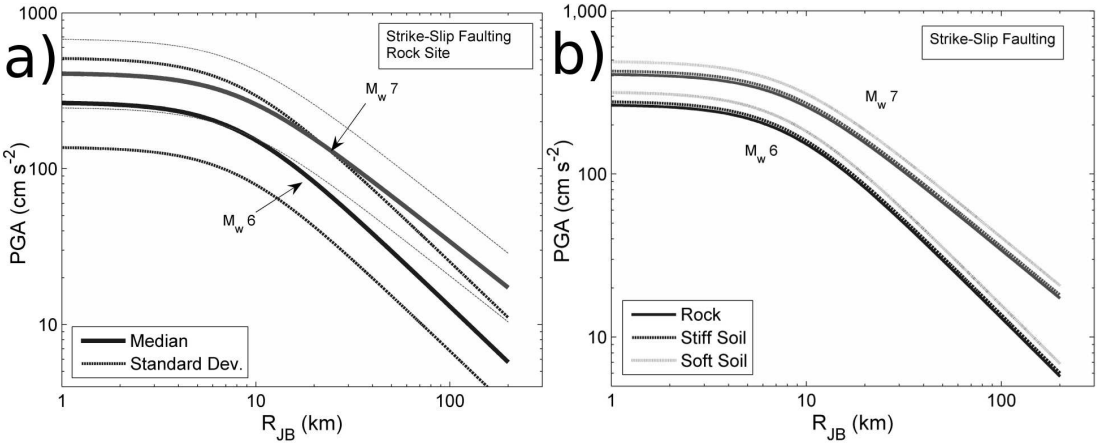

Figure 5.6: a) Comparison of 50^{th} and 84^{th} percentile attenuation using the Bommer *et al.* (2007) relation, b) Comparison of 50^{th} percentile attenuation on various rock sites

5.3 Attenuation Relations with Spectral Ordinates

For the purposes of site-specific seismic hazard applications it is often necessary to construct a uniform hazard spectrum (UHS). This is essentially a spectrum or accelerations with a probability of being exceeded over a specified period of time. To construct such a spectrum an empirical relation describing the attenuation of the spectra with distance is needed. Within the last fifteen years the importance of attenuation relations with spectral ordinates, for the purposes of site-specific hazard analysis has grown substantially. There are fewer available attenuation relations for this purpose. In selecting spectral attenuation relations, all the factors that influence the choice of PGA relations still apply. There are also several other factors to take into consideration. As noted by Cotton *et al.* (2006), the spectral range of the relation is an important consideration for engineering purposes. Accompanied with that is the recognition that the spectra filtering technique applied to the data set and the response of the accelerographs may both influence the results. Extracting strong motion spectra from analogue records is an expensive process and is often not done for small ground motions (Douglas, 2003). Consequently, there is often

a greater correlation between magnitudes and the distances of the records, with smaller earthquakes often being recorded only within the nearest 40 - 50 km. What follows is a description of the spectral ordinates of some of the attenuation relations described previously.

5.3.1 Ambraseys *et al.* (1996)

The strong motion records in this relation come mostly from free-field stations, with some from basements or ground floors of structures. Records from distances greater than the shortest distance to the nearest non-triggered accelerograph were also included so as to avoid error from malfunctioning equipment. No correction is made for instrument characteristics, as for many sites there is insufficient information to do so. For records of strong shaking, a bandpass filter between 0.20 and 20 Hz is applied.

A two-stage regression procedure is applied, for which the anelastic attenuation parameter is found to be insignificant after the first regression. As with the zero-period relation, soil conditions are characterised by two parameters: S_A and S_S, which are equal to one if the site is stiff soil or soft soil respectively, and 0 otherwise. Analysis of the residuals with magnitude and distance reveal no linear trends that are significant at the 5 % level. Spectral ordinates are damped at 5 % critical and span the range 0.10 s to 2.00 s. The function for both horizontal and vertical acceleration takes the form:

$$\log_{10}(a_i) = C_{1i} + C_{2i}M_W + C_{4i}\log_{10}(r_i) + C_{Ai}S_A + C_{Si}S_S + \sigma \tag{5.27}$$

For $0.1 \leq T_i\ (s) \leq 2.00$, where $r_i = \sqrt{R_{JB}^2 + h_{0i}^2}$. The spectra for several different earthquakes are shown in Figure 5.7.

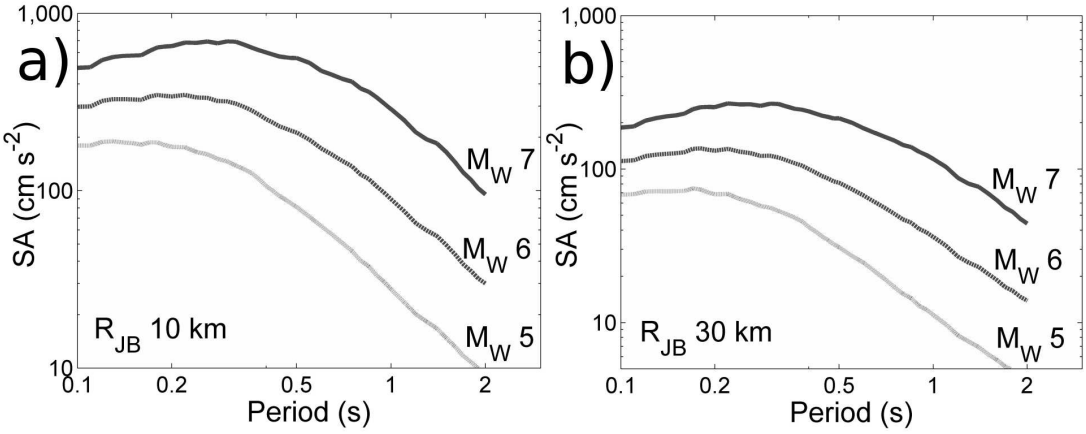

Figure 5.7: 50^{th} percentile 5 % damped for a Joyner-Boore distance of a) 10 km and b) 30 km, for a rock site and a normal fault using the Ambraseys *et al.* (1996) relation

The authors compare this relation to similar spectral attenuation relations from Italy and western North America. They find that acceleration for shorter periods tends to be greater in the near field for higher magnitudes and attenuates more slowly with dis-

tance. For smaller magnitude earthquakes the opposite appears to be true. Of all the relations this is suggested as being valid over the greatest range of magnitudes and distances. There is some error in the conversion between M_W and M_S, as well as in the determination of the Joyner-Boore distance. This attenuation relation has been widely applied in many examples over the last 10 years. As with the zero-period attenuation relation, earthquakes from Greece and Turkey are well represented, which suggests that it can be applied over the entire Aegean region.

5.3.2 Ambraseys *et al.* (2005a)

It has already been argued that the zero-period form of this relation that the dependence of sigma on magnitude can produce an undesirable effect in seismic hazard analysis. This also applies to the spectral ordinates.

Faulting mechanism for each earthquake is classified using the criteria of Frohlich and Apperson (1992). Here, earthquakes with their T-axis plunging greater than 50° are classified as thrust, with B- or P- axis plunging greater than 60° classified as strike-slip or normal, and everything else as odd.

Records were filtered above 23 Hz for analogue instruments, and at 50 Hz for digital instruments. The number of records available for regression at each period diminished rapidly at periods greater than 1s. The regression method used was a one-stage maximum likelihood method.

The correlation of scatter with magnitude is significant at the 5 % level, hence the magnitude-dependent uncertainty value. Accordingly, data were weighted by a linear function of M_W in the regression analysis.

Where this relation differs in shape from the Ambraseys *et al.* (1996) relations is in the dependence of decay rate on magnitude. This produces an additional scaling term of $(c_3 + c_4 M_W)$ for the distance decay. The equation then takes the functional form

$$\log_{10}(a_i) = c_{1i} + c_{2i}M_W + (c_{3i} + a_{4i}M_W)\log\left(\sqrt{R_{JBi}^2 + c_{5i}^2}\right) + \cdots$$
$$\cdots + c_{6i}S_S + c_{7i}S_A + c_{8i}F_N + c_{9i}F_T + c_{10i}F_O + \sigma_i P \tag{5.28}$$

for $0 \leq T_i(s) \leq 2.50$, where $\sigma_i = \left((b_{1i} + b_{2i}M_w)^2 + (b_{3i} + b_{4i}M_W)^2\right)$.

The 5 % damped response spectra are shown in Figure 5.8. For the vertical spectral ordinates not all sigma values are magnitude dependent.

When compared to similar attenuation relations, the Ambraseys *et al.* (2005a) relation

tends to amplify long period accelerations over soft soil sites to a greater extent. The authors express surprise that despite the addition of more parameters to characterise faulting, the standard deviations have not reduced to any extent. This is attributed to the dependence of the standard deviation on magnitude and the addition of more records over a greater magnitude and distance range. The latter has the impact of including records from greater distances, over which the variability in crustal anelastic attenuation increases.

Given the high proportion of records from Greece and Turkey, it is reasonable to assume that this attenuation relation can be readily applied to the entire Aegean region. Furthermore, by introducing parameters to characterise the earthquake fault type, this relation may be able to capture the variability in fault mechanism over small distances adequately, which is common in the Aegean region.

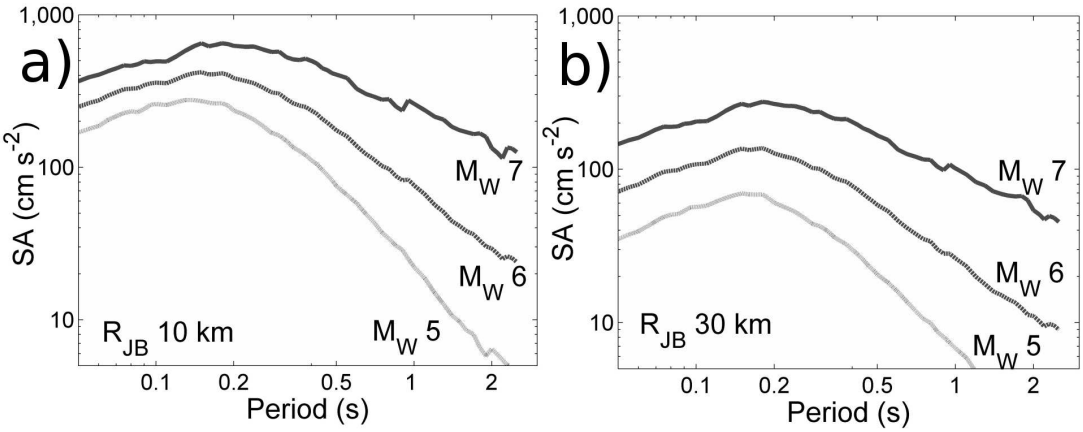

Figure 5.8: 50^{th} percentile 5 % damped response spectra for a Joyner-Boore distance of a) 10 km and b) 30 km, for a rock site and normal fault, using the attenuation relation of Ambraseys *et al.* (2005a)

The use of moment-magnitude will avoid error due to conversion between M_W and M_S, which was a problem with the Ambraseys *et al.* (1996) relation. Also, with the inclusion of the well recorded Izmit earthquake, a more reliable set of records for larger earthquakes in the Aegean is used; hence the upper bound of the magnitude range has risen. Unfortunately, the lower bound of the magnitude range has also risen, which means there is less confidence in the relation when extrapolating it to earthquakes in the range M_W 4 - 5. It has been shown by Bommer *et al.* (2007) that extrapolating the Ambraseys *et al.* (1996) relation to M_W 4 earthquakes drastically under-predicts the acceleration at shorter periods. The effect of the correlation of standard deviation with magnitude for smaller events will also be greater. Disaggregation of a simple PSHA using this relation shows that the hazard is controlled mostly by small earthquakes, which is contrary to the demands of the engineering structures (Musson, 2009)!

Although this relation passes the criteria for use in the Aegean region, a substantial degree of caution should be observed. This relation has yet to be heavily tested in practical

applications of PSHA, and its shortcomings have already been recognised. There is an argument that could be made for increasing the weighting of this relation, in a logic tree analysis, for higher magnitude events, and increasing the weighting of the Ambraseys *et al.* (1996) relation for lower magnitudes, owing to the fixed sigma value.

5.3.3 Danciu and Tselentis (2007)

This relation includes 30 spectral ordinates for ground acceleration and velocity, which span the spectral range $0.1 \leq T(s) \leq 4$. The attenuation equation assumes the following form, and the response spectra can be seen in Figure 5.9:

$$\log_{10}(a_i) = a_{i0} + b_i M_W + c_i \log_{10}\left(\sqrt{R^2 + h_i^2}\right) + e_i S + f_i F + \sigma_i P \tag{5.29}$$

The depth coefficient, h, is considered to be a fictitious depth, which is determined as a parameter within the regression. The applicable depth range of the data set considered is between 0 and 30 km.

The shape of the spectrum is much more strongly influenced by magnitude in this model than those of Ambraseys *et al.* (1996) and Ambraseys *et al.* (2005a). With larger magnitudes the greatest acceleration values shift from a peak in the 0.1 - 0.2 Hz range to 0.5 Hz. However, the substantial increase in acceleration towards the large end of the magnitude range suggests non-linearity at these magnitudes. This suggests that extrapolating this relation to even greater magnitudes would produce unrealistically high acceleration values in the near field, often in excess of 2 g.

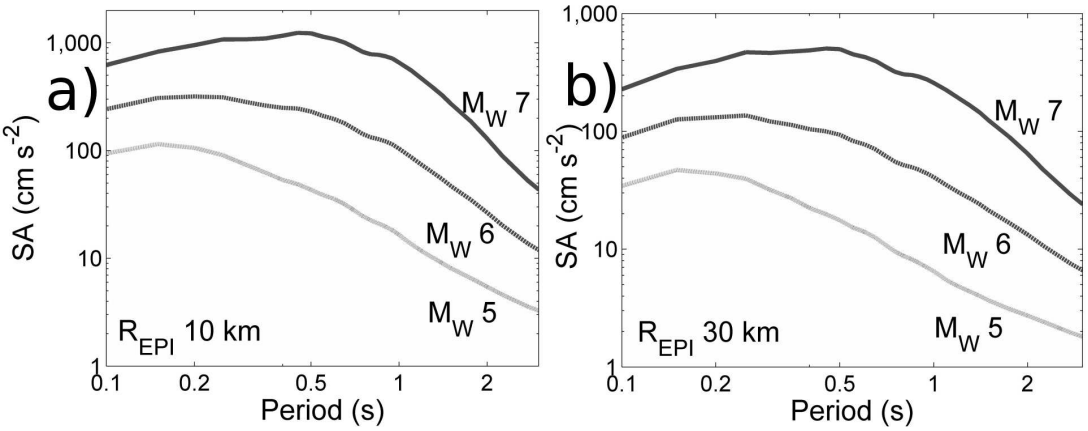

Figure 5.9: 50^{th} percentile 5 % damped response spectra for an epicentral distance of a) 10 km and b) 30 km, for an assumed normal fault and rock site, using the Danciu and Tselentis (2007) spectral attenuation relation

This attenuation relation is derived exclusively from Greek earthquakes. It is therefore reasonable to expect that it should be valid across all of Greece, and the extensional faulting regimes in Macedonia and Southern Bulgaria. There are no events in the data set representing the thrust earthquakes that are common along the Adriatic coast of Albania

and Montenegro. It is uncertain, whether this relation can be applied to those regions. The strike-slip events in the data set are mostly taken from the Ionian Islands, with only one earthquake occurring on the western end of the north Anatolian Fault in the North Aegean Sea. With no records pertaining to the dextral strike slip faulting found in north western Turkey, it is harder to justify applying this attenuation in that region. For the extensional faulting in western and south western Turkey, it may not be unreasonable to assume that anelastic attenuation due to the crustal structure is not dissimilar to that of central Greece.

5.3.4 Bommer *et al.* (2007)

This attenuation relation contains ordinates for ten spectral parameters, spanning the period range 0.05 s to 0.5 s. The number of parameters constrained is smaller than that of other spectral attenuation relations. This may be due to the extended magnitude range, for which the number of strong motion records with accurate accelerations at particular period is smaller. Consequently there are fewer long period ordinates over which a robust regression can be implemented.

As with the zero-period acceleration, both the inter- and intra-event aleatory variability for each of the spectral parameters also linearly decreases with increasing magnitude. Some discussion of this is given within the reference itself. Although it is argued that aleatory variability is heteroscedastic, the issue as to whether or not this is a manifestation of the limitations of the data set is not resolved. As with the previous European attenuation relations, soil amplification is modelled as a linear variable, correlating with NEHRP site class. The response spectra for this relation are shown in Figure 5.10.

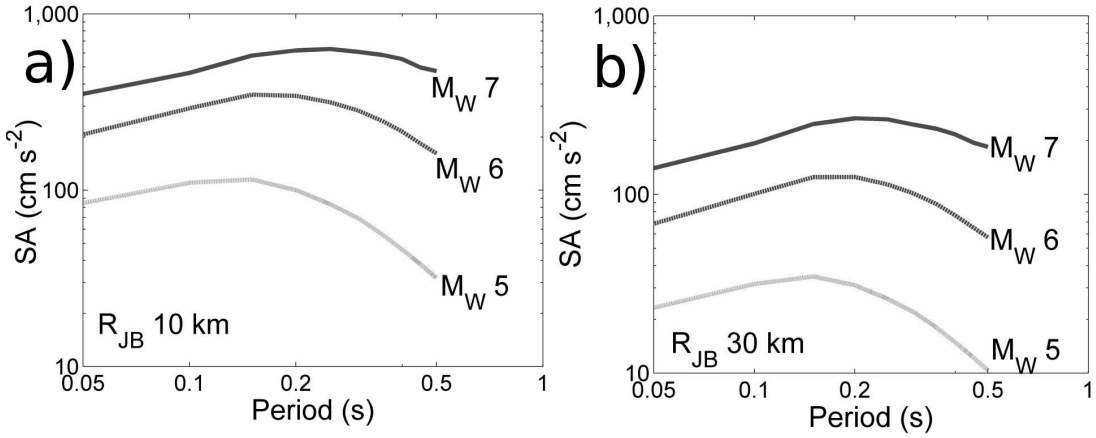

Figure 5.10: 50^{th} percentile 5 % damped response spectra for a Joyner-Boore distance of a) 10 km and b) 30 km, for an assumed normal fault and rock site, using the Bommer *et al.* (2007) attenuation relation

5.3.5 Selection of spectral attenuation relations

Of the four attenuation relations presented here for spectral acceleration, it is difficult to decide which relation should be used in a seismic hazard study across the whole Aegean region. The Ambraseys *et al.* (1996) model spans perhaps the widest range of magnitudes and distances. It also distinguishes between different site characteristics and has a fixed uncertainty for each period. The fact that it doesn't represent faulting characteristics makes it perhaps more suitable for hazard mapping in the Aegean, as the variability due to fault type is incorporated into the standard deviation of the relation. Unfortunately, the dependence on M_S means that an additional error enters the hazard analysis due to the conversion between M_W and M_S.

The relation of Ambraseys *et al.* (2005a) is dependent on M_W (directly determined from the original seismograms rather than via conversion), thus avoiding error from conversion of magnitude scales. There are, however, many concerns about the use of this relation. In particular, the linear relation between σ and M_W could prove to be a significant problem if the Monte Carlo simulation includes earthquakes in the range $4.0 \leq M_W \leq 4.5$. Given the large increase in uncertainty it is possible that hazard at a site will be controlled by small magnitude earthquakes. For larger earthquakes, however, this relation and that of Bommer *et al.* (2007) may have the most reliable data set, having included strong motion records from the 1999 Izmit and Duzce sequences. The Izmit earthquake was the largest in the 20th century Aegean catalogue, and most estimates for the regional maximum earthquake appear to be around M_W 8.5. Although it may not be desirable to extrapolate this attenuation to magnitudes greater than M_W 7.6, this may be the more realistic attenuation model from which to do so.

Many of the advantages associated with the Ambraseys *et al.* (2005a) relation are also applicable to the Danciu and Tselentis (2007) attenuation relation, in particular the dependence on M_W. Furthermore, the choice of epicentral distance metric in the Danciu and Tselentis (2007) relation is convenient for basic Monte Carlo seismic hazard, which simulates earthquakes as point sources with energy radiating from a focus. Where there are problems with this relation, however, is in the range of the strong motion data set used. The decision to use exclusively Greek earthquakes and none from the NAF in north-western Turkey limits the area over which this relation could be considered valid. Equally, the lack of strong motion records from larger earthquakes mean this relation should be extrapolated to magnitudes above M_W 7 only with caution.

The issue of extrapolating these attenuation relations, both spectral and PGA, beyond the magnitude and distance range of the data set used is an important one in seismic hazard analysis. As more strong motion records become available, it is clear that the linear magnitude scaling term that is used in most relations is not realistic. It is often the case that attenuation relations derived from moderate to large earthquakes under-predict the strong motion from small earthquakes (particularly at shorter periods), and those

derived from small to moderate earthquakes over-predict the strong motion from large earthquakes (Bommer *et al.*, 2007). The functional form of the Ambraseys *et al.* (2005a) attenuation relation may fit the data better by including a combined magnitude-distance scaling term and a linear correlation between magnitude and uncertainty. When this is applied in seismic hazard analysis the results may not be desirable, as the hazard then becomes controlled by small events.

5.4 Attenuation Relations using Arias Intensity

The attenuation relations presented here have focused on PGA and spectral acceleration. There are other ground motion parameters that may be of use in engineering applications. Some of the relations listed previously have also performed regressions for peak ground velocity (PGV) and Peak Ground Displacement (PGD) (Joyner and Boore, 1981; Theodulidis and Papazachos, 1992; Margaris *et al.*, 2001; Skarlatoudis *et al.*, 2003; Danciu and Tselentis, 2007). Alternatively, some recently published attenuation relations consider energy based ground motion parameters such as Arias Intensity (Travasarou *et al.*, 2003; Danciu and Tselentis, 2007; Stafford *et al.*, 2008a).

5.4.1 Danciu and Tselentis (2007)

This model has performed a regression for a variety of engineering parameters as well as spectral velocity. Considered further in this work is the relation for Arias Intensity. The attenuation relation has the same functional form as that for PGA. Here I_a is the sum of the two horizontal Arias Intensities in $cm\ s^{-1}$:

$$\log_{10}(I_a) = -2.663 + 1.125M_W - 2.332\log_{10}\left(\sqrt{R^2 + 13.092^2}\right) + \cdots \\ \cdots + 0.028S + 0.200F + 0.524P \tag{5.30}$$

Energy based parameters, such as Arias Intensity, are sometimes used in hazard analysis because they often show stronger correlations with damage than PGA, PGV or PGD. Many of these parameters are a function of amplitude and duration of strong motion, which is an important factor in determining damage to structures. Of these relations Arias Intensity and CAV have been most widely studied and their relation to damage understood. As a single parameter these values can be of particular use in hazard mapping.

5.4.2 Travasarou *et al.* (2003)

This is a worldwide attenuation relation exclusively using Arias Intensity (I_a). The data set contains 1208 strong motion records from 75 earthquakes worldwide, of which 11

earthquakes are from Greece or Turkey. The arithmetic mean of the two horizontal components is used as the parameter for regression, this is in contrast to Danciu and Tselentis (2007) who use the sum of the two horizontal components (equation 5.30). Earthquakes in the data set are evenly distributed across the magnitude range $4.7 \leq M_W \leq 7.6$, and rupture distance $0.1 \leq R_{RUP}\,(km) \leq 250$. This is represented by the parameters S_C and S_D, where S_C is 1 for stiff soil and 0 otherwise, and S_D is 1 for soft soil and 0 otherwise. Faulting is represented by two parameters F_N and F_R. F_N is 1 for normal faults, 0 otherwise, and F_T is 1 for thrust and oblique faults, 0 otherwise. The regression equation is given as follows:

$$\begin{aligned}\ln(I_a) = {} & 2.800 - 1.981\,(M_W - 6) + 20.72 \ln\left({M_W}/{6}\right) - 1.703 \ln\left(\sqrt{R_{rupt}^2 + 8.78^2}\right) + \cdots \\ & \cdots + (0.454 + 0.101\,(M_W - 6))\,S_C + (0.479 + 0.334\,(M_W - 6))\,S_D - \cdots \\ & \cdots - 0.166F_N + 0.512F_R + \sigma P \end{aligned} \tag{5.31}$$

where σ is the vectorial mean of the intra- (ξ) and inter-event (τ) errors, which are defined as:

$$\tau = 0.611 - 0.047\,(M_W - 4.7) \tag{5.32}$$

$$\xi = \begin{cases} 1.18 & I_a \leq 0.013 ms^{-1} \\ 1.18 - 0.106\,(\ln(I_a) - \ln(0.0132)) & 0.013 < I_a < 0.125 ms^{-1} \\ 0.94 & I_a \geq 0.125 ms^{-1} \end{cases} \tag{5.33}$$

It is quite clear that this attenuation has a different functional form than those seen previously, in particular that of Danciu and Tselentis (2007). The coefficients of this relation are determined via the random effects model approach. What is distinctive about this model is the non-linearity between Arias Intensity and magnitude. This is defined by the $\ln(M_W/6)$ term. The addition of this term is designed to reduce the influence of small near-field earthquakes on hazard. Typically small earthquakes are associated with a shorter duration of shaking, thus reducing the likely damage. Also of note is the inclusion of linear magnitude dependence on the site category. This too is different from other attenuation relations and, again, reflects the non-linearity of the duration of strong motion on soils.

Given the difference in function form, it is difficult to compare the Travasarou *et al.* (2003) to that of Danciu and Tselentis (2007). A comparison is shown for the 50^{th} percentile and 84^{th} percentile in Figures 5.11 a) and b) respectively. It is clear that there is reasonable agreement between the two relations (having adjusted for the difference in definition of horizontal I_a) for larger earthquakes, but that the Danciu and Tselentis (2007) relation predicts much lower values of I_a for smaller to moderate earthquakes. This can be attributed to the difference in magnitude scaling term, which is linear in Danciu and

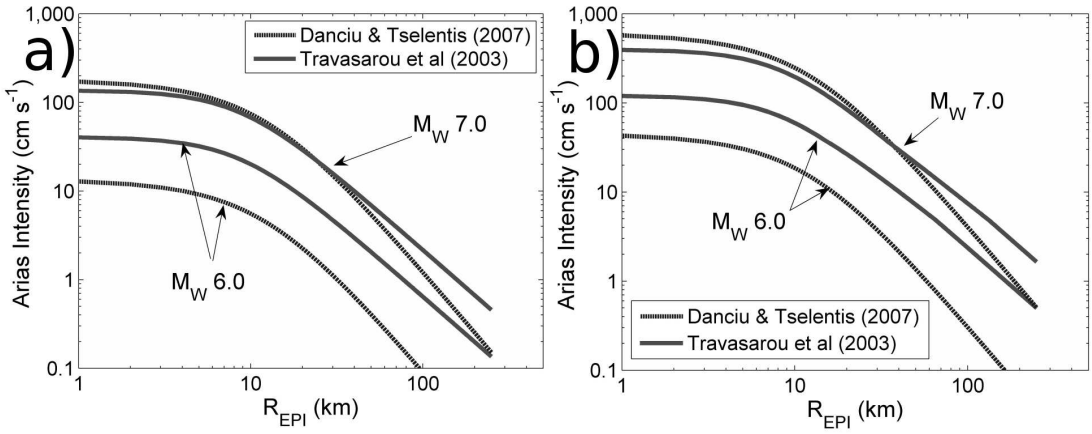

Figure 5.11: Comparison of the Travasarou *et al.* (2003) [solid line] and Danciu and Tselentis (2007) [dashed line] Arias Intensity attenuation functions for a "rock" site and a strike-slip fault. a) 50^{th} percentile motion and b) 84^{th} percentile motion

Tselentis (2007) and non-linear in Travasarou *et al.* (2003). It remains to be seen what impact the non-linear scaling of magnitude, standard deviation and site coefficients has on seismic hazard analysis undertaken using these relations. The Danciu and Tselentis (2007) Arias Intensity relation is preferred here, however, due to its emphasis on Greek earthquakes.

5.5 Next Generation Attenuation Relations

In 2006 and 2007 the field of strong ground motion modeling underwent an apparent transition with the publication of the "Next Generation Attenuation" (NGA) relations. This was a project established by the California based Pacific Earthquake Engineering Research Centre, with a view to a coherent development of attenuation relations using a single consistent data set. Five attenuation relations were produced: Boore and Atkinson (2007) - BA07, Campbell and Bozorgnia (2007) - CB07, Chiou and Youngs (2007) - CY07, Idriss (2007) - ID07 and Abrahamson and Silva (2007) - AS07. The data set used contains 3552 strong motion records from 175 different earthquakes from various active tectonic regions. There is a distinct bias toward western California earthquakes. However, the most abundant set of strong motion recordings come from the 1999 Chi-Chi earthquake in Taiwan (M_W 7.6) and its aftershocks, which are often considered separately from the remainder of the catalogue in the regressions.

It is typical for attenuation relations for any region to be updated every decade or so as new strong motion recordings and numerical techniques are used. As can be seen in this chapter, with each generation of relations there has been an increase in complexity. For example the addition of a site term, a faulting term, separation of inter- and intra-event errors. The NGA relations take this even further with the inclusion of non-linear and mixed linear site effects, hanging wall effects and non-linear magnitude scaling. Fur-

thermore, many of the relations have included various new parameters defining both the earthquake and the site (e.g. depth to top of the rupture, rake, dip, depth to V_S = 2.5 km/s interface etc). In addition to the introduction of new parameters, many scaling functions have now increased in complexity.

Whilst the increased complexity and number of parameters may help to reduce aleatory variability, though that in itself is debatable, this does not come without cost. In particular, outside of regions such as the western United States, where faults and well constrained and money is available for extensive site investigation, it is difficult to correctly constrain all the necessary parameters. This is especially true for the mapping applications, where knowledge of faulting and site class can vary spatially. The Aegean region is such a place where faulting and site parameters are ill constrained. Fewer than half of the strong motion recording sites have accurate estimates of V_{s30}, the fundamental site parameter common to all NGA relations (Stafford *et al.*, 2008b). European and Greece-specific relations have tended to characterize site amplification using a broad definition of the soil condition from NEHRP class. Given that NEHRP site classification is based upon, among other things, V_{s30}, comparison between NGA and European models is feasible, assuming a reasonable value of V_{s30} for each site class. Campbell and Bozorgnia (2006) suggest $V_{s30} = 1130\ m\ s^{-1}$ for Rock (Class B) sites, $V_{s30} = 560\ m\ s^{-1}$ for Stiff Soil (Class C) sites and $V_{s30} = 270\ m\ s^{-1}$ for Soft Soil (Class D) sites.

A full treatment of each of the NGA relations is well beyond the scope of this thesis. Following the example of Stafford *et al.* (2008b) the Boore and Atkinson (2007) relation will be considered in detail, and implemented alongside existing European relations. This particular relation is arguably the most practical relation for application in Europe as it requires the fewest input parameters. It is noted, however, that this relation does underestimate short-period spectral amplitudes for small earthquakes at short distances ($R_{JB} < 20$ km).

5.5.1 Boore and Atkinson (2007)

This attenuation model provides regression parameters for PGA, PGV and Spectral Acceleration 22 spectral ordinates in the range 0.01s to 10s. It is suggested as being valid for a magnitude range of $5.0 \leq M_W \leq 8.0$ and over a Joyner - Boore distance of 0 to 200 km. The functional form is described below:

$$\ln(Y) = F_M(M_W) + F_D(R_{JB}, M_W) + F_S(V_{s30}, R_{JB}, M_W) + \varepsilon\sigma_T \tag{5.34}$$

where Y is the ground motion in g for PGA and SA, and cm s^{-1} for PGV, σ_T is the root mean square of intra and inter-event variability.

$$F_D(R_{JB}, M_W) = [c_1 + c_2(M_w - M_{ref})] \ln\left(\frac{R}{R_{ref}}\right) + c_3(R - R_{ref}) \tag{5.35}$$

c_1, c_2 and c_3 are period-dependent coefficients, M_{ref} = 4.5 and R_{ref} = 1.0. $R = \sqrt{R_{JB}^2 + h^2}$ where h is a period-dependent coefficient.

$$F_M(M_W) = \begin{cases} e_1 U + e_2 SS + e_3 NS + e_4 RS + e_5 (M_W - M_h) + \cdots & \\ \cdots + e_6 (M_W - M_h)^2 & \text{if } M \leq M_h, \\ e_1 U + e_2 SS + e_3 NS + e_4 RS + e_7 (M_W - M_h) & \text{if } M > M_h. \end{cases} \quad (5.36)$$

where e_1, e_2, ... e_7 and M_h are period-dependent coefficients. U = 1 if fault type is unknown, 0 otherwise, S_S = 1 if strike-slip, 0 otherwise, N_S= 1 if normal faulting, 0 otherwise, and R_S= 1 if reverse faulting, 0 otherwise.

$$F_S = b_{LIN} \ln\left(\frac{V_{S30}}{760}\right) + F_{NL} \quad (5.37)$$

b_{LIN} is a period dependent coefficient. F_{NL} is defined thus

$$F_{NL} = \begin{cases} b_{NL} \ln(0.6) & pga4nl \leq 0.03 \\ b_{NL} \ln(0.006) + c [\ln(pga4nl/0.03)]^2 + \cdots & \\ \cdots + d [\ln(pga4nl/0.03)]^3 & 0.03 < pga4nl \leq 0.09 \\ b_{NL} \ln(pga4nl/0.1) & 0.09 < pga4nl \end{cases} \quad (5.38)$$

pga4nl is the estimate of 50^{th} percentile PGA on a reference site (V_{s30} = 760 m s^{-1}).

$$c = (3\Delta y - b_{NL}\Delta x)/\Delta x \quad (5.39)$$

$$d = -(2\Delta y - b_{NL}\Delta x)/\Delta x^3 \quad (5.40)$$

where $\Delta x = \ln(0.09/0.03)$ and $\Delta y = b_{NL} \ln(0.09)$. Finally:

$$b_{NL} = \begin{cases} b_1 & V_{s30} \leq 180 \\ (b_1 - b_2) \ln(V_{s30}/300) / \ln(180/300) + b_2 & 180 \leq V_{s30} \leq 300 \\ b_2 \ln(V_{s30}/760) / \ln(300/760) & 300 < V_{s30} < 760 \\ 0.0 & 760 \geq V_{s30} \end{cases} \quad (5.41)$$

where b_1 and b_2 are period-dependent coefficients.

It is clear that the Boore and Atkinson (2007) relation is substantially more complex than the predecessors. Much of this complexity is in the non-linear site amplification. On a reference rock site ($V_{s30} = 760\ m\ s^{-1}$) the formulae are comparable to other relations for Europe. The soil non-linearity function is a development of the non-linearity model of Choi and Stewart (2005). This functional form requires a mixture of period-dependent and period-independent coefficients.

The expected magnitude range of this attenuation relation is $5 \leq M_W \leq 8$, over

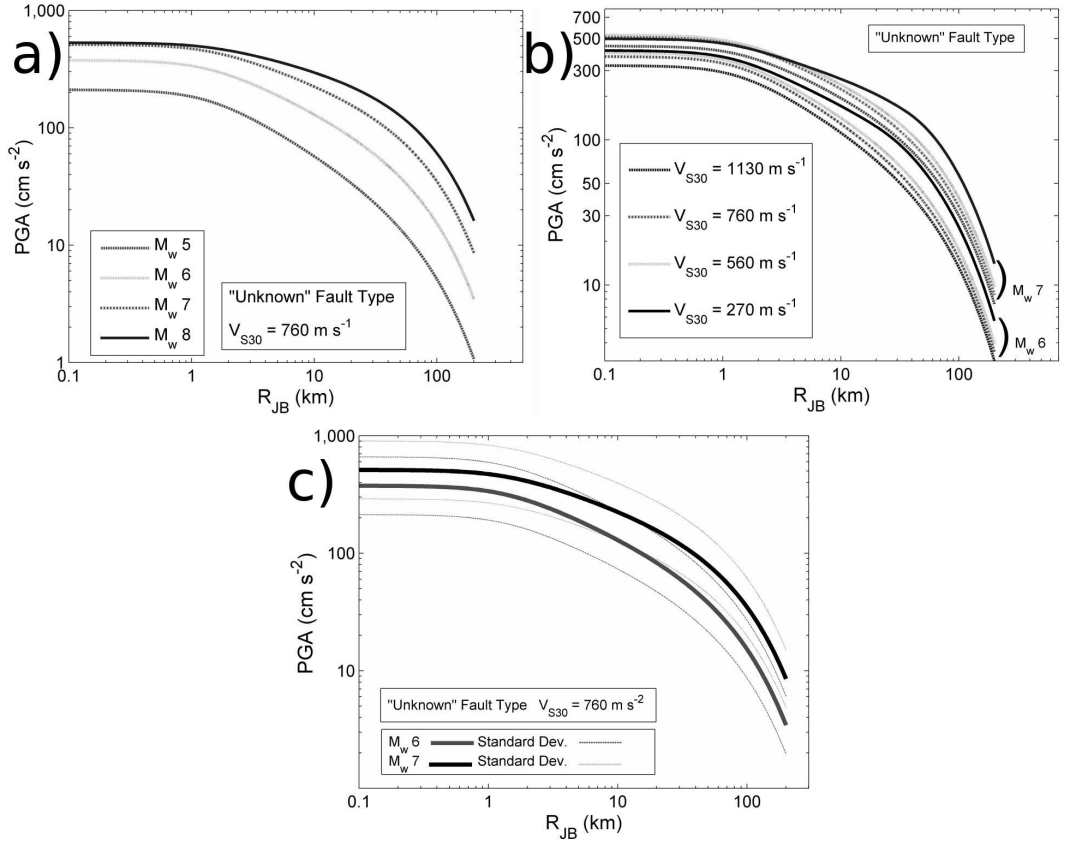

Figure 5.12: PGA attenuation using the Boore and Atkinson (2007) attenuation relation. a) Attenuation on a rock site with an "Unknown" fault type. b) Comparison of PGA attenuation on different site types, c) Comparison of 50^{th} and 84^{th} percentile PGA attenuation for a rock site and "Unknown" fault type.

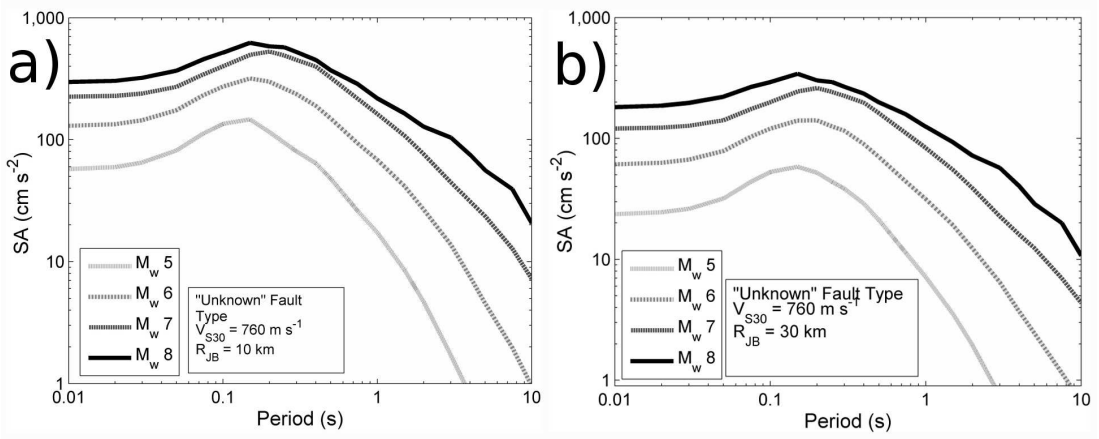

Figure 5.13: 50^{th} percentile 5 % damped response spectra for a Joyner-Boore distance of a) 10 km and b) 30 km, for an assumed rock site and "Unknown" fault type, using the Boore and Atkinson (2007) spectral attenuation relation

a Joyner-Boore distance range of 0 - 200 km. Site conditions range from $180 \leq V_{s30}\ (m\ s^{-1}) \leq 1300$, which spans the NEHRP site classes B, C and D. Only extremely soft saturated soils (Classes E and F) and very hard rock sites (Class A) are outside of this range. This attenuation relation also covers a large spectral range from 0.01s to 10s, which allows for application to large heavily engineered structures. This has come at a cost of poorer spectral resolution in the period range 0.01s to 0.5s, which may be of more importance for hazard to residential properties. The PGA relation is plotted in Figure 5.12, and the response spectra in Figure 5.13.

Perhaps the most notable feature of this relation is small difference in the acceleration (both spectral and PGA) between M_W 7 and M_W 8 in the near field. This is also observed in the spectral peaks in the 0.1 - 1s range for these magnitudes. This saturation of acceleration in the near field may help to prevent anomalously high hazard owing to poor constraint of the attenuation from large earthquakes at short distances.

5.6 Selection of Attenuation Relations for Use in the Aegean region: Qualitative Basis

5.6.1 Comparison of Attenuation Relations

The issue of selecting attenuation relations for use in seismic hazard analysis is an ongoing area of debate. Using the criteria of Cotton *et al.* (2006) it is possible to reject several candidate models immediately. Whilst the Joyner and Boore (1981) relation may have been considered appropriate for the Aegean region, the absence of Aegean earthquakes is cause for concern, especially given that other relations are available. This relation also lacks a parameter for soil type. The absence of an uncertainty parameter in the Makropoulos and Burton (1985b) relation is the main reason for rejection. Also, the dependence on M_S rather than M_W introduces an additional error when applied in the Monte Carlo simulations here.

The Theodulidis and Papazachos (1992) relation is also immediately rejected. Although it is derived from a catalogue of Greek earthquakes it exhibits several problems, not least of all the dependence on M_S. The main reason for rejection is the mischaracterisation of soil types. Clearly if there is a known error in the data set, and other models are available, there would appear to be no reason to consider this model further. This is in spite of the suggested modification suggested by Theodulidis (1998), which appears arbitrary and is hard to justify when considering this relation alongside others.

Two more relations can be rejected on the basis that they have been superseded by subsequent relations. These are Ambraseys (1995) and Margaris *et al.* (2001). Although there are some differences in the relations, the Ambraseys (1995) relation is superseded by Ambraseys *et al.* (1996) and Ambraseys *et al.* (2005a), both of which included spectral

ordinates. Margaris *et al.* (2001) never appeared in peer reviewed literature. However, a near identical data set is used in the development of the Skarlatoudis *et al.* (2003) relation. Thus the Skarlatoudis *et al.* (2003) relation is preferred.

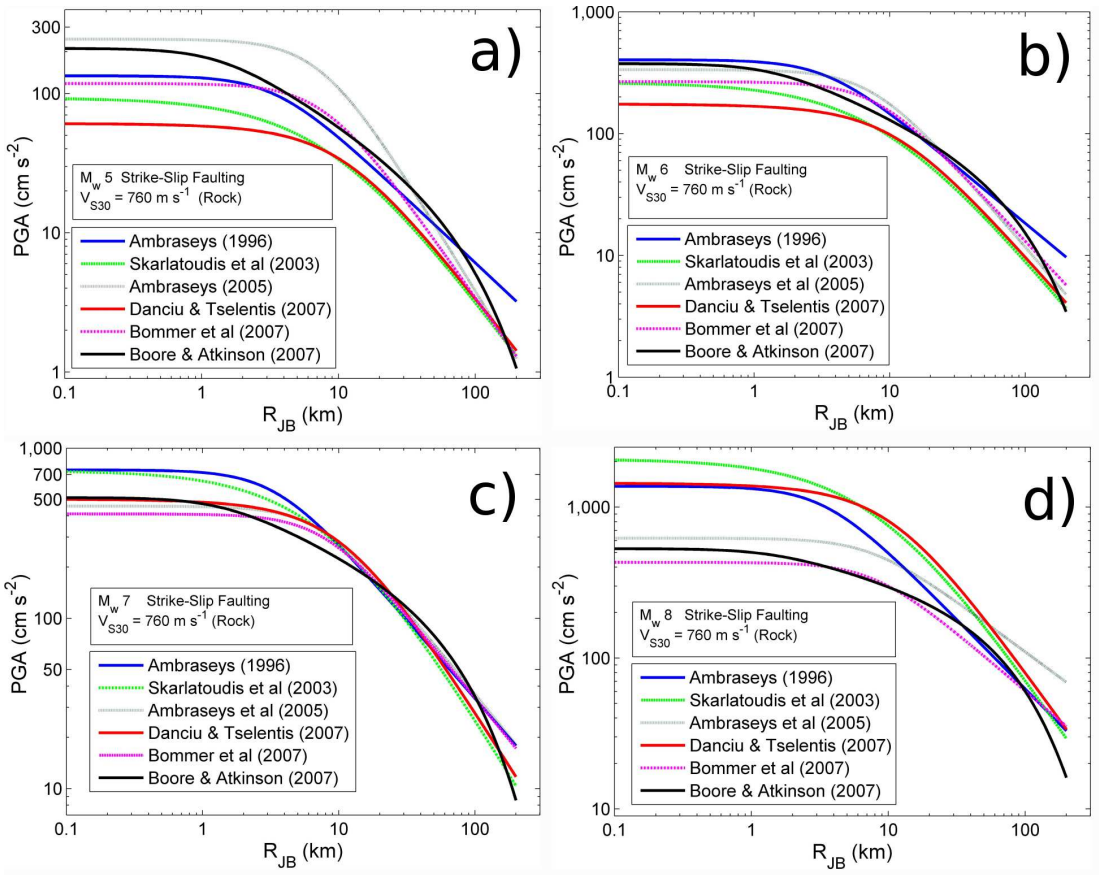

Figure 5.14: Comparison of the six candidate PGA attenuation relations for use in the Aegean region for an assumed strike-slip fault and rock site ($V_{S30} = 760m\ s^{-1}$), over a Joyner-Boore distance range of 0.1 to 200 km. a) M_W = 5, b) M_W = 6, c) M_W = 7 and d) M_W = 8

The attenuation relations that are not rejected are Ambraseys *et al.* (1996) [Am96], Skarlatoudis *et al.* (2003) [Sk03], Ambraseys *et al.* (2005a) [Am05], Danciu and Tselentis (2007) [DT07], Bommer *et al.* (2007) [Bm07] and Boore and Atkinson (2007) [BA07]. Of these relations, all contain both PGA and spectral ordinates except for Sk03. The magnitude scale used in most of these relations is M_W, the exception being Am96. European and global relations use Joyner-Boore distance as the distance parameter whereas Greek-specific relations use epicentral distance. This arises due to the issue of poorly constrained seismic sources in the Aegean region, which has already been discussed.

Although Ambraseys *et al.* (1996) may be superseded by Ambraseys *et al.* (2005a), and Bommer *et al.* (2007), it is retained for further analysis despite the fact that this is the only remaining relation to use M_S as a magnitude parameter. This is because, unlike the latter Europe-wide attenuation relations, aleatory variability is homoscedastic over all spectral ordinates. It is therefore useful to compare the results with those latter relations for which

aleatory variability is heteroscedastic.

If the PGA attenuation models for all six relations are compared some interesting trends emerge (Figure 5.14). In particular, there is a consistent contrast in the way in which Greek specific and Europe-wide relations behave at different magnitudes. For smaller magnitudes the two Greek specific relations (SK03 and DT07) predict lower accelerations across most of the distance range. There is reasonable agreement between all the relations for M_W 6 - M_W 7, especially over distances greater than 10 km. At M_W 8 the Greek relations and Am96 predict near field accelerations in excess of 1g, whereas Am05, Bm07 and BA07 suggest near-field accelerations in the 600 to 700 cm s^{-2} range. This contrast is mostly expected because of the extrapolation of the Greek attenuation relations well beyond their applicable range. In fact only BA07 can be considered feasible for attenuation from great earthquakes ($M_W \geq 8.0$), although extrapolation beyond M_W 8 is not advised.

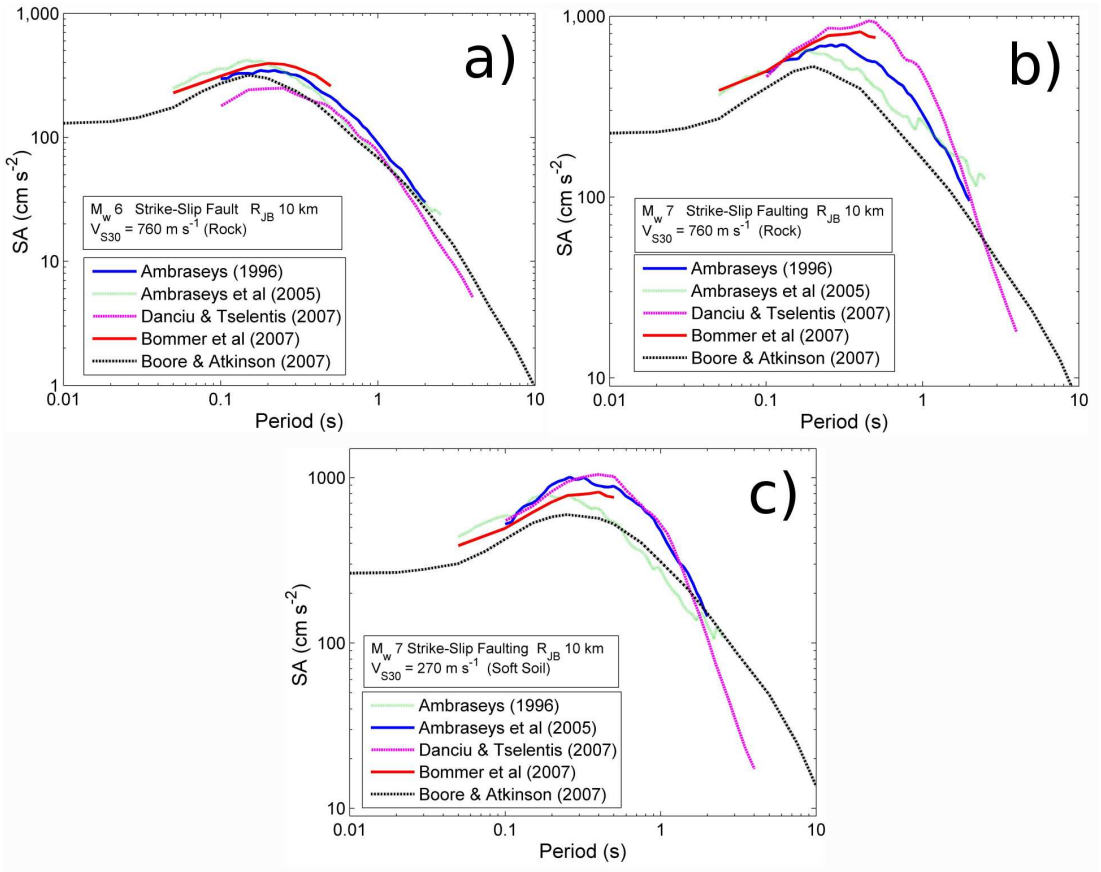

Figure 5.15: Comparison of response spectra for five candidate spectral attenuation relations for use in the Aegean. Joyner-Boore distance is 10 km and strike slip faulting is assumed. a) M_W = 6 on a rock site ($V_{S30} = 760m\ s^{-1}$), b) M_W = 7 on a rock site and c) M_W = 7 on a soft soil site ($V_{S30} = 270m\ s^{-1}$)

A comparison of the five spectral relations is shown in Figure 5.15. There is reasonable agreement in spectral acceleration between the models for moderate sized earthquakes (M_W 5 - 6.5), with only DT07 producing appreciable lower accelerations in the 0.1 s - 0.3

s range. At higher magnitudes there is good agreement amongst all but the BA07 models in the 0.1s - 0.3s range (on rock), but then diverging considerably at longer periods. DT07 produces the highest spectral accelerations at a rock site for M_W 7, which is important since this is the only relation that is extrapolated beyond its derived limits. This makes the agreement between Am05, DT07 and Bm07 for M_W 7 on soft soil all the more surprising. The BA07 relation produces lower accelerations at M_W 7 than for the other relations, but has perhaps the most significant increase in longer period accelerations on soft soil sites. This is likely due to the non-linear scaling of ground motion on soil, which is included in BA07, but not for the other relations. The difference between DT07 and BA07 at the longest periods (T $>$ 2 s) for large earthquakes on soft soil suggests that the DT07 may under-predict long period accelerations if extrapolated to larger events.

5.6.2 Conversion from R_{EPI} to R_{JB}

It should also be noted that although PGA is plotted against R_{JB} in Figures 5.14 and 5.15, SK03 and DT07 were converted from epicentral distance to Joyner-Boore distance. This is not a straightforward matter, and consideration must be given as to how to convert between different distance scales. Were faults in the Aegean well-constrained, and the location of the hypocentre on the fault plane known, conversion between epicentral and Joyner-Boore distance could be achieved via simple geometry. Using appropriate fault scaling relations (Wells and Coppersmith, 1994) and moment tensors, fault geometry could be estimated with reasonable accuracy. Within the European Strong Motion database (Ambraseys *et al.*, 2004), nearly half of the strong motion records do not give Joyner-Boore distances, only epicentral and hypocentral distance. Since fault geometries are not constrained in the Monte Carlo simulation of earthquakes in the Aegean region, geometric conversion is not possible.

In lieu of geometric conversion, the remaining option is to convert via empirical relations. This brings with it an additional source of error, especially for larger events, but may be the most computationally efficient approach. Several empirical relations are suggested (Montaldo *et al.*, 2005):

$$R_{JB} = -3.5525 + 0.8845 R_{EPI} \qquad \text{Ambraseys and Bommer (1991)} \tag{5.42}$$

$$R_{JB} = -5.0497 + 0.9433 R_{EPI} \qquad \text{Ambraseys } \textit{et al.} \text{ (2004)} \tag{5.43}$$

$$R_{JB} = \left[\left(R_{EPI}^2 + 25 \right) 10^{0.566 - 0.114 M_S} - 5.8^2 \right]^{1/2} \qquad \text{Montaldo } \textit{et al.} \text{ (2005)} \tag{5.44}$$

Equation 5.42 is suggested for European data, equation 5.43 and 5.44 for Italian data. A more computationally complex method is also suggested by Scherbaum *et al.* (2004b), who relate epicentral distances to Joyner-Boore distances using the following relation:

$$R_{JB} = R_{EPI} + \varepsilon_{JBepi} \tag{5.45}$$

where ε_{JBepi} is a gamma distributed random variable with density:

$$f(x;\alpha,\beta) = \begin{cases} (\beta^{\alpha}\Gamma(\alpha))^{-1} x^{\alpha-1} \exp\left(-x/\beta\right) & \text{for} \quad x > 0 \\ 0 & \text{for} \quad x \leq 0 \end{cases} \tag{5.46}$$

This conversion allows for uncertainty in hypocentral location on the fault plane, but at considerable computational cost. Furthermore, the conversion parameters required for calculation of the gamma function (as given in Scherbaum *et al.* (2004b)) are magnitude-dependent and are derived using the empirical fault scaling relations of Wells and Coppersmith (1994). This introduces further uncertainty into this conversion method.

Given the additional complexity of modelling using a gamma distribution, the simpler empirical relations (Equations 5.42 - 5.44) are preferred. These conversions produce a similar scaling to the Scherbaum *et al.* (2004b) method for moderate sized earthquakes (Montaldo *et al.*, 2005). It is also found that epicentral distance is approximately equal to 5 km when Joyner-Boore distance is equal to zero. This corresponds to the distance validity range of the Sk03 and DT07 relations, suggesting reasonable equivalence in the near-field. In the next section the preferred empirical conversions will be compared within the likelihood analysis of ground motion fit. In the Monte Carlo simulations the Ambraseys and Bommer (1991) relation (Equation 5.42) is used, as it was in Figures 5.14 and 5.15.

5.7 Applicability of Attenuation Relations to the Aegean Region: Quantitative Basis

5.7.1 Measuring fit of an attenuation model

In section 5.6 six attenuation relations for PGA (five with spectral ordinates) were identified as being applicable to the Aegean region (Am96, Sk03, Am05, DT07, Bm07, BA07). Each of these attenuation relations has different properties and different ranges of magnitude and distance over which they can be applied. The question therefore arises as to which of these relations represent, with any degree of accuracy, the real attenuation of strong motion in the Aegean region.

It is possible that several of the relations indicated may model strong motion in the Aegean region in a satisfactory manner. If this is the case then these relations will need to be incorporated into the epistemic uncertainty of the seismic hazard analysis. This may typically be done using a logic tree or via further Monte Carlo methods. Regardless of which method is used, each attenuation model (or branch of the logic tree) must be apportioned an appropriate weighting; a stronger weighting indicating a better fitting model. Considerable investigation into how to assign weightings to the strong motion attenuation model has been undertaken by several authors (Scherbaum *et al.*, 2004a; Bommer

et al., 2005; Cotton *et al.*, 2006).

A quantitative method of assessing the validity of an attenuation relation within a region is suggested by Scherbaum *et al.* (2004a), and implemented subsequently in the Pyrenees (Drouet *et al.*, 2007), Central Italy (Bindi *et al.*, 2006), the French Antilles (Douglas *et al.*, 2006) and Europe-wide by Stafford *et al.* (2008b). The Scherbaum *et al.* (2004a) method is a maximum likelihood calculation that compares observed strong motion with model predictions.

If Z is the normalised ground motion, defined as $Z = \frac{x-\mu}{\sigma}$ where μ is the 50^{th} percentile predicted ground motion, x the observed ground motion and σ the aleatory variability of the ground motion model, then the goodness of fit of a model to the observed data is defined as:

$$LH\left(|Z|\right) = Erf\left(\frac{|Z|}{\sqrt{2}}, \infty\right) = \frac{2}{\sqrt{2\pi}} \int_{|Z|}^{\infty} \exp\left(\frac{-z^2}{2}\right) dz \tag{5.47}$$

Here LH is the likelihood value, which reaches a maximum of 1 for Z = 0 (observation is equal to the mean value of the model). If the samples are drawn from a standard normal distribution, the likelihood values are uniformly distributed between 0 and 1. The median likelihood value is used as an indication of fit, with values around 0.5 indicating strongest fits between observed and modelled data. Although the likelihood value is a novel indicator of the goodness of fit of observed strong motion data to strong motion modelled from attenuation relations, other information can still be useful. In particular the mean, median and standard deviation of the normalised ground motions still need to be considered.

Using these statistics to describe the fit of an attenuation relation to strong motion data, it is possible to make some decisions as to how valid the said attenuation relations are to a specific region. Scherbaum *et al.* (2004a) recommend a ranking scheme, illustrated in Table 5.2.

The classification scheme illustrated here can be used as the basis for designing a weighting scheme to model epistemic uncertainty in the ground motion relation. Weightings can be apportioned according to the model classification. It is appropriate to exclude entirely attenuation relations of Class D, and possibly even Class C depending on the number of good fitting models.

5.7.2 Strong Motion Data

Observed strong motion data are taken from the European Strong Motion Database Ambraseys *et al.* (2004). Earthquakes and strong motion records from Greece, Turkey (west of $32^{\circ}E$), Albania, Macedonia and Southern Bulgaria were selected. Strong motion records

Table 5.2: Scherbaum *et al.* (2004a) classification criteria for fit of strong motion attenuation relations to observed strong motion data.

Class	LH_{MEDIAN}	$\|E(Z)\|$	Z_σ	Validity
A	≥ 0.4	≤ 0.25	≤ 1.125	Excellent Fit - Applicable to region
B	≥ 0.3	≤ 0.5	≤ 1.25	Good Fit - Applicable, but some error
C	≥ 0.2	≤ 0.75	≤ 1.5	Moderate - Applicable, but undesirable
D	< 0.2	> 0.75	> 1.5	Unacceptable - Should not be used

on NEHRP class E sites were excluded. This produces a total of 146 records from 37 different earthquakes, including the 1999 Izmit and Duzce earthquakes, and several major aftershocks. The full list of records and their notable parameters is given in Appendix B.

Where spectral ordinates are tested, the spectral accelerations are damped at 5 %. The observed spectral range is 0.04 s to 2.5 s, sampling 66 specific periods of oscillation. Comparison is only made for horizontal spectral acceleration. Vertical acceleration and horizontal displacement and velocity are not considered. The strong motion records indicate that a Butterworth low cut filter of 0.20 Hz has been applied, and Butterworth high cut filter, cosine tapered between 23 Hz and 25 Hz, has also been applied. Sampling interval is 0.005 s.

Accelerations are corrected to correspond to the appropriate strong ground motion definition for each relation. Consequently, for Am96 and Am05 the larger of the two horizontal components of motion is used. For Bm07 and BA07 the geometric mean is used. Boore and Atkinson (2007) specify that it is the orientation independent geometric mean (Boore *et al.*, 2006) that is preferred. DT07 use the arithmetic mean of the two components, whilst Sk03 do not specify how acceleration is characterised; geometric mean is presumed. Variation in faulting is taken into consideration since all fault mechanisms are present in the strong motion database.

Differences in site coefficient require some discussion. One of the greatest shortcomings of the European Strong Motion Database is the incomplete characterisation of the recording sites. Estimates of V_{s30} are given for 85 of the 146 strong motion records. Elsewhere only the Eurocode site classifications are given. For those relations that require NEHRP site classification, conversion is made from the Eurocode classification. For BA07 V_{s30} is required. Where V_{s30} values are missing the following values are defaulted to: NEHRP D = V_{s30} 300 $m\ s^{-1}$; NEHRP C = V_{s30} 570 $m\ s^{-1}$; NEHRP B = V_{s30} 760 $m\ s^{-1}$. This is a reasonable approximation using the intermediate values in each range (Stafford *et al.*, 2008b).

For those records lacking a R_{JB} value, conversion is made from R_{EPI} using equations 36 (R_{JB}AB), 37 (R_{JB}Amb) and 38 (R_{JB}Mon). This allows for comparison of the three empirical equations listed in section 5.6.2 and their performance in fitting strong ground motion. Where R_{JB} is given for a strong motion record, the value is used.

Table 5.3: Principal factors in the determination of PGA attenuation classification for Greece. Z_{50} refers to the median normalised value. Numbers in bold indicate Class A level, those in italics indicate class C level, underscore indicates class D (not visible here, but seen in spectral parameters - Appendix B). Otherwise the level is class B

Attenuation Relation	Magnitude	Distance	E(Z)	Z_{50}	Z_σ	LH_{50}	Class
Am96	M_S	R_{JB}AB	-0.48	-0.42	*1.30*	0.35	C
Am96	M_S	R_{JB}Amb	-0.47	-0.45	*1.30*	0.37	C
Am96	M_S	R_{JB}Mon	-0.33	-0.26	*1.35*	0.35	C
Sk03	M_W	R_{EPI}	**0.14**	**0.17**	1.20	0.39	B
Am05	M_W	R_{JB}AB	*-0.62*	-0.46	*1.44*	**0.41**	C
Am05	M_W	R_{JB}Amb	*-0.62*	-0.46	*1.44*	**0.42**	C
Am05	M_W	R_{JB}Mon	-0.47	-0.25	*1.49*	**0.42**	C
DT07	M_W	R_{EPI}	**0.05**	**0.05**	1.22	0.38	B
Bm07	M_W	R_{JB}AB	-0.32	**-0.23**	*1.42*	**0.43**	C
Bm07	M_W	R_{JB}Amb	-0.32	**-0.23**	*1.42*	**0.43**	C
Bm07	M_W	R_{JB}Mon	**-0.19**	**-0.05**	*1.47*	**0.40**	C
BA07	M_W	R_{JB}AB	**0.05**	**-0.02**	**1.12**	**0.53**	A
BA07	M_W	R_{JB}Amb	**0.05**	**0.01**	1.13	**0.53**	B
BA07	M_W	R_{JB}Mon	**0.21**	**0.17**	1.16	**0.53**	B

Inspection of the model fit for PGA reveals some interesting results (Table 5.3). The only relation to achieve 'A' classification is BA07, which is the only relation constructed using principally data from outside the European and Mid-East region. The three Europe-wide relations (Am96, Am05 and Bm07) all perform poorly compared to the Aegean-specific relations. This is particularly surprising since Greek and Turkish strong motion records account for more than half of the strong motion records in the database. Nevertheless the LH_{50} values for Am05 and Bm07 still exceed 0.4. These two relations are classified as class 'C' on the basis of high Z_σ. This arises from the heteroscedastic aleatory variability associated with these two relations. There is still a tendency for the European relations to under predict strong ground motion for the Aegean data set. There is no consistent improvement in fit when R_{EPI} to R_{JB} conversion methods are used; hence it is not unreasonable to proceed with the Ambraseys and Bommer (1991) conversion in future use.

5.7.3 Fit of PGA Attenuation Relations

The two Greece-specific relations perform reasonably well, which is also surprising given the prevalence of records from the Izmit and Duzce events in the data set. The improvement in performance over the European relations may suggest that a degree of bias is being introduced in the distance conversion process, but it is unclear to what extent this is the case. The LH_{50} values for Sk03 and DT07 are lower than those of Am05 and Bm07, despite the lower absolute mean of normalised residuals. This would suggest that the principal difference between European and Greece-specific relations arises from the differences in modelling of aleatory variability. It may also suggest that the heteroscedastic model of variability is not necessarily the most appropriate for the Aegean. This is also

supported by the goodness of fit of the BA07 model, which models aleatory variability as homoscedastic.

5.7.4 Fit of Spectral Attenuation Relations

Analysis of the model fit for spectral ordinates is given in Figure 5.16, with the specific values given in Appendix B. Rather than interpolating the spectra for each attenuation relation onto the ordinates found in the database, only ordinates common to each relation and the database are tested. Here there is a considerable difference in the models when compared to PGA. It is immediately clear that no model maintains category 'A' across all periods for all three indices of fitness quality. Of the relations compared Bm07, DT07 and BA07 remain within category 'B' across almost all applicable periods when E(Z) and LH_{50} are considered. Indeed it is Bm07 that often has the lowest $|E(Z)|$ at each spectral ordinate. It is noticeable that BA07 performs most strongly in the 0.06 - 0.12s period range. At longer periods, however, there is a substantial increase in Z_σ. The opposite is true for Bm07, which has a large Z_σ for short periods, diminishing rapidly as period increased. This may again be due to the heteroscedastic aleatory variability, which increases at lower magnitudes; hence the increase in variability at shorter periods. For much of the 0.1 - 0.5 s range Z_σ falls into category D, despite the good fit in the other indices.

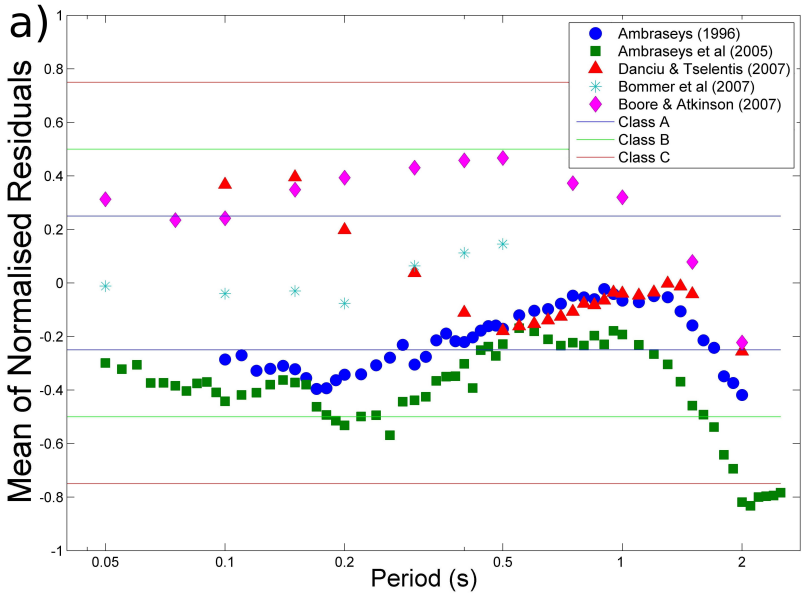

Figure 5.16: a) Mean of Normalised Residuals

The Am96 and Am05 relations vary considerably in the quality of fit across their expansive spectral range. As with Bm07, Z_σ is very large in the 0.1 - 0.5 s range, falling well into category D. There is a general improvement in fit for both Am96 and Am05 in the 0.5 - 1.5 s ranges, before diminishing again at T $>$ 1.5 s. As with Bm07, the high Z_σ is likely due to

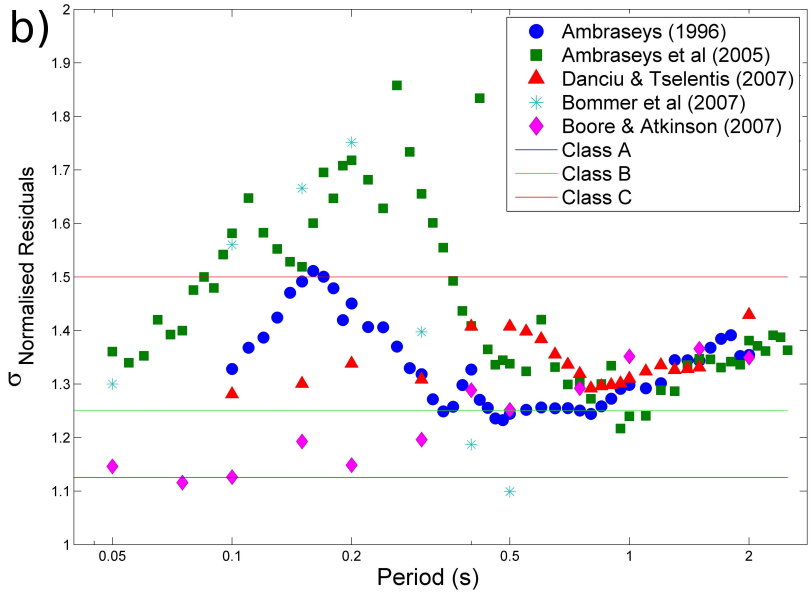

Figure 5.16: b) Standard Deviation of Normalised Residuals

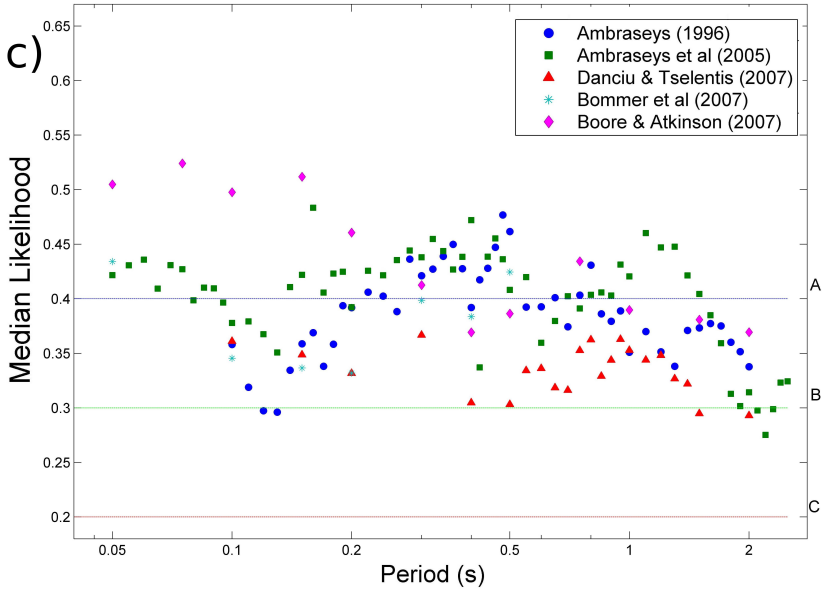

Figure 5.16: Comparison of attenuation model validity for spectral acceleration. a) E(Z), b) Z_σ and c) LH_{50}

the heteroscedastic aleatory variability. The Am05 model is also the only one to consider spectral period in the 2 - 2.5 s range. The fit is generally poor in that range, however. This suggests there is little to be gained in including Am05 simply for its greater spectral range and longer periods.

Having considered each of the attenuation models and their spectral ranges, which should be applied for seismic hazard analysis? Within this, another issue arises as to how to balance to quality of PGA fit against those of spectral parameters. For the purposes of constructing hazard maps it may be the case that a model is preferred on the basis of PGA fit, rather than on those of its spectral ordinates. This assumes that PGA and SA are practically exclusive, which is not necessarily the case in practise since hazard spectra are anchored using zero-period acceleration

Considering PGA first, it is clear that BA07 has the best overall fit. Even when allowing for error in conversion from epicentral distance to Joyner-Boore distance, this would still be the strongest weighted model. DT07 and Sk03 both give category 'B' fits. In the absence of spectral ordinates, however, there is good reason to select DT07 for use and not Sk03. When the average effective depth and the variable depth versions of Sk03 are compared, the variable depth gives a poorer fit, although still within the same category 'B'. This may be due to errors in the depth estimates for some events in the catalogue. On this basis there is little reason to keep both DT07 and Sk03, in which case DT07 is preferred.

Perhaps the most significant decision is how to approach the use of the European relations (Am96, Am05 and Bm07). Firstly it is clear that Am96 offers little improvement in fit to justify the use of M_S alongside M_W. This is despite being the only one of the three relations to have homoscedastic aleatory variability. In terms of PGA there is a slight improvement of fit in the Bm07 relation as opposed to Am05, although both remain in the same category on the basis of LH_{50}. This may give grounds for affording Bm07 a slightly stronger weighting for PGA.

The differences in fit are less clear when considering spectral acceleration. No one attenuation model achieves class A or B across the entire sample spectral range. The one that remains in these categories for the greatest range is BA07, which remains in class A and B for the range 0 - 0.4 s, stabilising as category C above 0.5 s. DT07 remains in classes A and B over its entire spectral range for E(Z) and LH_{50}, but class C for Z_σ. Similarly, Am96 performs reasonably well for E(Z) and LH_{50}, and remains a consistent class C for Z_σ. Bm07 and Am05 both fall into class D for Z_σ in the 0.1 - 0.5 s range, but perform better than many other relations outside this range.

5.7.5 Selection of Attenuation Relations for use in Seismic Hazard Analysis

When selecting the spectral attenuation model using this technique, an important consideration is the context in which it will be used. For a site specific analysis the choice of relation may depend on the structure being considered and the fundamental response period. For large engineered structures the fit of the model in longer periods may be of greater importance than for shorter periods. Conversely, damage to most residential properties may be more strongly correlated with acceleration at shorter periods. For seismic hazard mapping, these considerations, whilst not trivial, may not be as extensively analysed. Again, it may depend on the purpose for which the hazard map is being created. If looking at likely damage to properties then an intermediate spectral value or duration dependent parameter (Arias Intensity) may have the strongest correlation to damage (Cabañas *et al.*, 1997).

In this study it is general variation in seismic hazard being considered, without a specific context of application. This means that performance of the model over a particular spectral range, or density of sampling within said range, may not be as important as the overall fit of the model over its full spectral extent. Under these conditions BA07 would seem the preferred candidate as it fits to an 'acceptable' level over most of its range and spans the largest spectral range, including periods up to 10 s. BA07 is therefore given the strongest weighting. In terms of spectra, both DT07 and Am96 give reasonable fits over a range, although Am96 contains a greater number of spectral ordinates. As with PGA, however, Am96 will contain a greater degree of error owing to discrepancies in magnitude and distance parameters. Despite Am96 sampling from a greater magnitude range, DT07 is preferred.

There is little difference between the datasets used in the development of Am05 and Bm07, albeit with smaller earthquakes included in Bm07. Over the magnitude and distance range being considered here there is little to distinguish between Am05 and Bm07. Both give high Z_σ within the same range. Bm07 does give a better fit when considering the mean of the normalised residuals (class 'A' across its entire range). Despite the fact that Bm07 contains fewer spectral ordinates it is preferred over the Am05 model.

5.8 Proposed Weighting Scheme for Analysis of Epistemic Uncertainty

The justification given for selection of models leads to the selection of four attenuation relations for consideration in an analysis of epistemic uncertainty. In order of strength of fit, these are: BA07, DT07, Bm07 and Am05. The decision therefore has to be made as to how to develop a weighting scheme for use in seismic hazard analysis. Four schemes are

Table 5.4: Proposed weighting schemes for analysis of epistemic uncertainty in modelling strong ground motion in the Aegean region

Model	PGA Class	EXP	EXP+	LIN	SOC
BA07	A	0.5	0.7	0.4	0.25
DT07	B	0.25	0.2	0.3	0.25
Bm07	C	0.15	0.08	0.2	0.25
Am05	C	0.1	0.02	0.1	0.25

proposed here for comparison, as shown in Table 5.4:

The first two schemes (EXP and EXP+) represent an exponential decrease in weighting with model class. EXP+ exaggerates the shape of the exponential, thus apportioning a greater weight to BA07 and lower weights to Bm07 and Am05. The LINEAR model represents a linear decay in weight with decreasing class. The final model, labelled SOCRATES ("All we know is that we know nothing"), is borrowed from the labelling system given in (Scherbaum *et al.*, 2005). With the exception of the SOCRATES model, Bm07 is given a greater weighting than Am05. These weighting schemes will be compared in the seismic hazard analyses performed in Chapter 6.

There are many issues to consider in the implementation of a weighting scheme for the purposes of an analysis of epistemic uncertainty (be it logic tree or Monte Carlo). Arguably the greatest complication is the difference in functional form between many of these equations. The difference between BA07 and the others is the primary example. All the relations have M_W as the preferred distance metric. However, only DT07 uses epicentral distance, which means conversion needs to be made for the other metrics. The preferred conversion equations lack an aleatory uncertainty term, which means that this variability cannot be incorporated into the aleatory variability of the attenuation relation. This is a substantial error as the aleatory variability can increase by a factor of 1.4 at short distances ($R_{JB} < 10\ km$) (Bommer *et al.*, 2005). Similarly, the style-of-faulting term differs slightly for each relation. With so many different definitions of fault class in the contributing data sets it is not possible to incorporate style-of-faulting conversion into the epistemic uncertainty analysis.

Perhaps the greatest distinction between the several relations is the different definition of horizontal ground motion used. Bm07 and Am05 use the larger component of PGA, whilst DT07 use the geometric mean and BA07 the orientation-independent geometric mean. Investigation into the relative scaling of different ground motion definitions (Bommer *et al.*, 2005) reveals that for PGA the 50^{th} percentile acceleration may be a factor of 1.14 greater for larger component relations than for geometric mean. This factor decreases to unity for longer periods of acceleration. The scaling factor for aleatory variability is close to unity. This may imply that PGA predicted by the Am05 and Bm07 relations may need to be scaled down for reasonable comparison with the DT07 and BA07. The difference in scaling between geometric mean and period-independent geometric mean is unknown

and assumed to be unity.

The assignment of weights in an analysis of epistemic uncertainty remains a subjective process, even with knowledge of the fit of each of these relations to Greek strong motion data. The four schemes proposed here represent four separate lines of thought: bias to the best fitting relations (EXP), very strong bias to the best fitting relation (EXP+), bias only by rank fit for each relation (LINEAR) and no bias at all (SOCRATES). The last category is interesting to explore as it assumes no difference in fit of each of these attenuation relations, and hence no judgement. The other categories represent a subjective judgement. Within this judgement there are still several questions that could be asked. Firstly, should attenuation relations with seemingly trivial weightings (e.g. Am05 in the EXP+ category) even be considered? The corollary to that being that if one attenuation is so strongly weighted, such as BA07 in the EXP+ scheme, then why consider any of the others? It should be made clear that these weighting schemes are implemented purely to explore the differences in bias to the strongest fitting relations. A seemingly infinite number of alternative schemes could also be considered, though the differences manifest in the hazard maps may be trivial.

A surprising result in this analysis is that the BA07 relation produced the best fit to the observed Greek data despite being derived from a worldwide set of earthquakes. This is in preference to the DT07 relation, which is derived exclusively from Greek earthquakes. The most likely reason for this better fit is that the strong motion test data set contained records from the Izmit and Duzce earthquakes and several aftershocks. These records were not found in the Greek dataset used to derive the DT07 relation.

5.8.1 Conclusions

This chapter has shown how it is possible to identify a number of candidate attenuation models from a larger selection, which could be applied to the Aegean region. By analysing, quantitatively, the fit of the models to observed strong motion data, selection can be made. In the absence of a single outstanding model, the relative fits can be used as a basis for weighting in an analysis of epistemic uncertainty. It is noteworthy that all of the candidate models have been developed within the last three years. This lends credence to the belief that strong motion models must be updated regularly and that older ones may not be appropriate for modelling within a region. Also of interest is the strong fit of the NGA relation (BA07), in this case in preference of European and Greece-specific relations. The NGA relation is the only one to be taken from a global dataset, with a bias towards Western U.S. and Taiwan (owing to the prevalence of records from the 1999 Chi-Chi event). This would suggest that NGA relations may be applicable to Europe, or at least to the Eastern Mediterranean. This observation is borne out by more detailed analysis in other research areas (Campbell and Bozorgnia, 2006; Stafford *et al.*, 2008b). This is an encouraging development as the NGA relations extend to higher magnitudes

than European or Greek relations, allowing for better constraint of strong motion from the largest events.

The issue of homoscedastic verses heteroscedastic variability is not resolved within the seismological community in general. The heteroscedastic variability found in Am05 and Bm07 is likely to be the reason for the high Z_σ values in the likelihood fits, and consequently lower rankings. Given the fit of Bm07 when other indices are compared, a case could be made for stronger weighting. Alternatively, it is also possible that the likelihood method of Scherbaum *et al.* (2004a) is not applicable to relations with hetero- and homoscedastic variability simultaneously. These are issues that will develop in seismological research in the future.

Chapter 6

Time-Independent Seismic Hazard Analysis for the Aegean Region

6.1 Introduction

The previous chapters have addressed the various inputs needed for seismic hazard analysis, and have identified models that are applicable to the Aegean region. Various alternatives models have been considered and attempts made to quantify their respective suitability to the Aegean region. It is obvious that, on the basis of previous analyses, no single model can be considered a definitive interpretation of seismogenic source or of strong motion attenuation. A comprehensive seismic hazard analysis for the Aegean region must therefore compare these alternative schemes. This will be done by considering each of these models and combinations of models separately (hereafter termed a "single model" approach) and then collectively as part of an analysis of epistemic uncertainty.

To recap, briefly, the results of the previous chapters, there are three main areas of possible epistemic uncertainty in the seismic hazard analysis: the seismogenic source model, the magnitude recurrence model and the strong motion attenuation model. In Chapter 2 the preferred magnitude recurrence model (the bounded Gutenberg and Richter (1944) model) is described. Consideration has been given to alternative magnitude recurrence models and arguments against their usage, both physical and statistical, are presented.

A collection of seismogenic source models have been presented in chapters 3 and 4. These models include existing source zonations (PP2000 and PZ1990), a novel zone model based on this author's interpretation of Aegean seismotectonics (WT2006) and models derived from the application of K-means cluster analysis to the spatial distribution of hypocentres and ruptures in the Aegean. Of the latter model set, the 27-zone (K27) and 29-zone (K29) ruptures partitions will be considered further.

The synthetic catalogues produced here are created using a hybrid of uniform source zones and distributed seismicity. Earthquakes simulated within a source zone are distributed uniformly across the zone, whilst outside the source zones seismicity is simulated by random re-sampling of the observed epicentres. For the PP2000, K27 and K29 models only a very small proportion of events will occur outside the source zones. For the re-sampled seismicity, b-value and its uncertainty are taken from the regional parameters.

The last area of considerable uncertainty is found in the selection of attenuation relation. Six attenuation models were identified in chapter 5 as being applicable to seismic hazard in the Aegean region. Hazard maps are produces for each of the six relations. The Sk03 model will not be considered in site-specific applications.

All the seismic hazard analyses have been implemented using original code in Matlab^{TM}. These programs can all run on desktop versions of Matlab, and have been done so here. Some of the larger Monte Carlo routines may be better run on high-performance systems, particularly those producing synthetic catalogues of considerable duration or with a low minimum magnitude. Matlab's intrinsic pseudorandom number generators for uniformly distributed and normally distributed random numbers are used. The uniform random number generator, based on the Marsaglia and Zaman (1991) algorithm, generates random numbers in the range 2^{-53} to $1 - 2^{-53}$, producing 2^{1492} numbers before repetition. The normal random number generator uses the Marsaglia ziggurat algorithm (Marsaglia and Tsang, 2000), which has a much smaller period of 2^{64} before repetition. To prevent repetition, each synthetic catalogue produced is seeded using the CPU clock.

The seismic hazard analyses presented in this chapter take two forms: site-specific and mapped. Further discussion on the elements necessary to create a map will be given in due course. Eight sites have been selected as test locations for the site-specific analyses. These are listed in Table 6.1, and shown in Figure 6.1. These cities are well spaced across a variety of different seismotectonic regimes.

Table 6.1: Eight cities considered for site specific seismic hazard analysis. V_{S30} values are taken from Wald and Allen (2007)

City	Longitude	Latitude	V_{S30} $\left(ms^{-1}\right)$	NEHRP Site Class
Argostoli	20.491	38.173	405	C
Athens	23.718	37.974	600	C
Heraklion	25.134	35.338	500	C
Istanbul	29.006	41.066	500	C
Izmir	27.145	38.433	300	D
Rhodos	28.222	36.443	550	C
Thessaloniki	22.973	40.625	250	D
Tirane	19.832	41.331	450	C

The hazard curves that will be produced for each site are intended as a means of com-

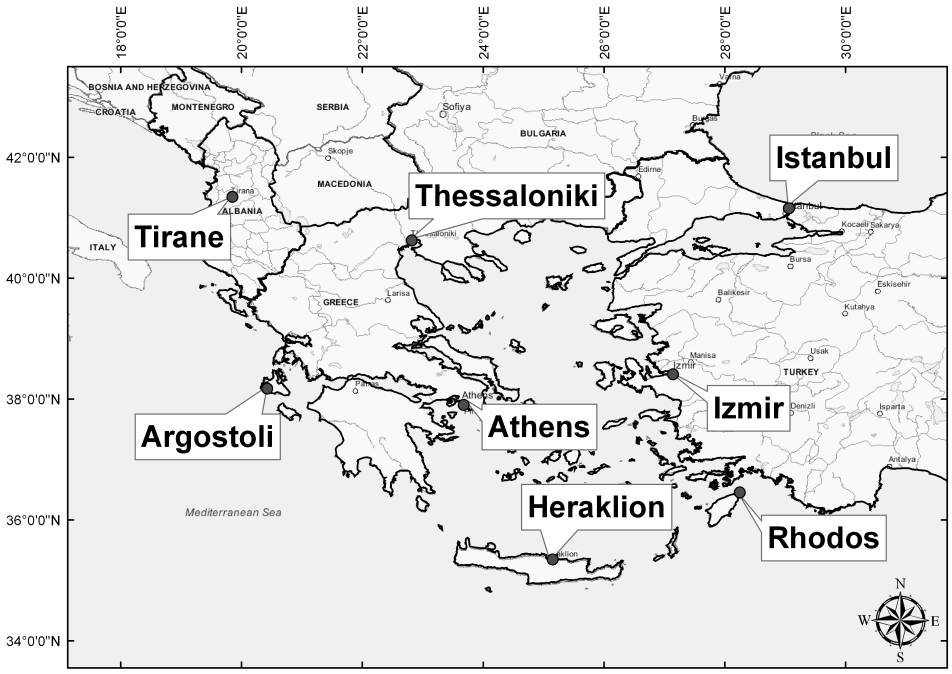

Figure 6.1: Eight test sites considered for the site-specific analysis application of the Monte Carlo seismic hazard programs

paring hazard across a range of probabilities of recurrence, rather than at specified probability levels. Uniform hazard spectra are presented for these sites for some models. Analyses at these sites are used for guidance and illustration of the Monte Carlo method and should not be used as a basis for engineering design in the cities. Further detailed site investigation would be needed for these hazard curves and hazard spectra to reach the level required by current building codes.

6.2 Operation of the Monte Carlo Procedure

6.2.1 Hazard Calculation

The basic premise of a hazard analysis via Monte Carlo simulation is explained in chapter 1. Seismic hazard (the ground motion with a P % probability of being exceeded in T years) is calculated directly using stochastically simulated earthquakes. The ground motion (a) with a fractional probability (P) of exceeded level (a_0) in a period of T years is given by:

$$P\left(a > a_0\right) = 1 - \exp\left(\frac{-T}{T_0}\sum_i H\left[a_i - a_o\right]\right) \tag{6.1}$$

where T_0 is the duration of the synthetic catalogue and H the Heaviside step function ($H\left[x\right] = 1$ if $x \geq 0$, and 0 otherwise). The summation term simply describes the total number of events in the synthetic earthquake catalogue that produce a ground acceleration exceeding a_0. This equation assumes that the occurrence of ground motion exceed-

ing a_0 is a stationary Poisson process.

Ground motion at a site is calculated directly from earthquakes in the synthetic catalogue. If these values are then ranked into descending order of strength it is possible to create a curve of annual probability (T = 1) of being exceeded simply by plotting P against H for ($H = 1,\ 2\ \ldots N_{T0}$), where N_{T0} is the number of earthquakes in the synthetic catalogue contributing to hazard at a site. For a specified probability level (e.g. 10 % probability of being exceeded in 50 years), equation 6.1 is re-arranged to find the number of times (N_H) the ground motion is greater than a_0 in the catalogue T_0. The ground motion for that probability of being exceeded is the N_H^{th} value of the ranked ground motions.

The seismic hazard software allows for two options for hazard calculation. To produce a seismic hazard map, it is necessary to specify the meridian and parallel bounds of the region being mapped, as well as the grid spacing (all in decimal degrees). Alternatively the software can produce hazard curves concurrently for a small number of sites. This requires the input of the location and soil condition for each site, and will produce a curve of annual probability of being exceeded.

This method of calculating hazard from synthetic catalogues is based on that of Ebel and Kafka (1999). Here, ground motions for a site are ranked directly, irrespective of year, and hazard calculated using equation 6.1. This approach differs from the alternative of Musson (1999b) and Giardini *et al.* (2004), who calculate probabilities by ranking the annual extreme ground motions and retrieving annual probabilities based on the rank directly (i.e. the annual probability of 10^{-3} is identified as the 1,000th annual event). Both approaches should give similar results over a sufficiently long synthetic catalogue.

The synthetic catalogue must be of sufficient duration to allow for repeated recurrence of the largest earthquakes in the region. Musson (1999b) suggests at least 1,000 times longer than the time period being considered in the hazard calculation, and ideally more than 1,000 times longer than the return period of the hazard. This is particularly important when considering low probability occurrences. The historical earthquake catalogue for the Aegean identifies three great earthquakes ($M_W \geq 8.0$) in the past 2000 years. This would suggest that a catalogue would need to be at least 100,000 years in duration to allow for a robust estimation of the hazard.

Such a considerable duration of synthetic catalogue comes at great computational cost, especially in high seismicity regions such as the Aegean. A synthetic catalogue of 10^7 years duration requires approximately 1 GB memory. Calculations using data sets of this size may exceed the available RAM on a desktop computer. It may be necessary, therefore, to extract the hazard values from a concatenation of multiple synthetic catalogues of a shorter duration.

6.2.2 Characterisation of time in the synthetic catalogues

A crucial assumption in the simulation of synthetic catalogues is that seismicity is a stationary Poisson process. Discussion of seismicity rate can be found in chapter 2. Here, regional seismicity rate is used to constrain the number of earthquakes simulated in the synthetic catalogues. The number of earthquakes occurring in each zone in then assigned in proportion to the observed seismicity rate for each zone relative to the regional rate.

Following the method of Ebel and Kafka (1999) an incremental time counter method is implemented. The time counter is based on the observed inter-event time distribution for region-wide seismicity as described in Figure 6.2. The time distribution can be created using one of two ways, the selection of which is left to the user. The first is by random sampling from a lognormal distribution, with mean and standard deviation calculated from the inter-event times in the observed catalogue. The second method samples inter-event times directly from the observed catalogue. If $M_{MIN} \geq M_C$ for the whole catalogue, it is necessary to restrict the set observed inter-event times to the period of the catalogue for which $M_{MIN} \geq M_C$. For the Aegean region a minimum magnitude range of $4.8 \leq M_{MIN} \leq 5.2$ can be explored comfortably.

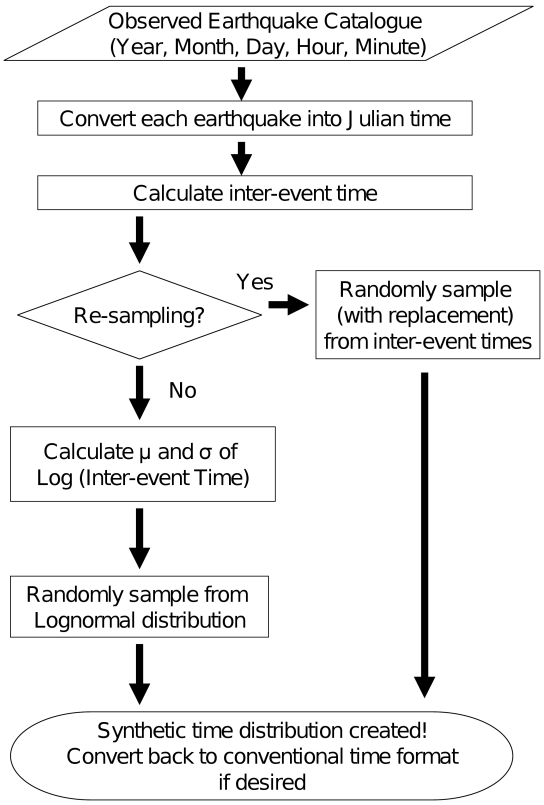

Figure 6.2: Flowchart describing the characterisation of time in synthetic catalogues

Incorporating time characterisation, here using Julian time, creates a simple framework from which non-Poissonian events can be later integrated (chapter 7). By defining time as a single real number, the synthetic catalogues can contain the same information as the

observed earthquake catalogue without a large increase in computational memory.

6.2.3 Creating the spatial distribution of seismicity

The implementation of the Monte Carlo procedure in this investigation utilises a hybrid of uniform seismogenic zones and spatially distributed seismicity. Earthquakes simulated inside a uniform zone are created by random sampling from a uniform distribution within the zone. Although many of the source models extend across the entire Aegean region, not all observed seismicity falls within a defined source zone. Since this seismicity may contribute to seismic hazard at a site, it cannot be overlooked in the analysis. The spatial distribution of seismicity outside of the uniform source zones is simulated by random re-sampling (with replacement) of the observed extra-zonal epicentres.

It has been well established that the use of uniform source zones can create an obvious ridge effect in hazard maps (Bender and Perkins, 1987). This may occur where a highly active seismic zone abuts a lower activity zone or the background seismicity. If the seismogenic source zone is not clearly defined by strong geological features, the appearance of ridges would appear to be unrealistic (Musson and Henni, 2001). To reduce, but not necessarily avoid, the effects of hard zone boundaries, measures must be taken to soften the zones. The Monte Carlo technique can do this by a two-step spatial sampling process. In the first step epicentres are sampled from within the uniform zone (or the extra-zonal seismicity) in the manner described previously. Another point is then sampled from a 2D Gaussian function with a mean position located on the previously sampled epicentre. The new point then becomes the final epicentre used in the catalogue.

This smoothing approach will soften the zone boundaries, without altering the seismicity rates for each zone in the manner that a non-zoned spatial smoothing approach might. The crucial parameter that controls the degree of smoothing is σ in the Gaussian function. A larger σ will increase the amount of smoothing, thus increasing the risk of obscuring zones of tightly constrained high activity. Equally, a low σ may not alter the spatial distribution sufficiently to allow for uncertainty in the observed epicentral location. Here σ is set to 0.1° ($\approx 11\ km$). This value is chosen for two reasons. Firstly, synthetic catalogues and hazard maps with different values of σ were compared, visually, with observed seismicity in the Aegean region and the delineation of seismic source zones. Of these maps $\sigma = 0.1^\circ$ represented the best compromise between replicating observed seismicity and distribution across the source zones. The second reason for selecting this value is that it represents the typical uncertainty in the instrumentally determined epicentral location for earthquakes in the pre-ISC Aegean catalogue (1900 - 1963). Earlier events, whose locations were inferred from macroseismic observations, had a slightly greater uncertainty in location than those that were instrumentally recorded (Makropoulos and Burton, 1981). Location uncertainty can vary considerably in epicentres derived from early instrumental records, owing to variability in network coverage.

6.2.4 Magnitude Limits

Much of the discussion of the issues in determining magnitude limits for the synthetic catalogues can be found in Chapter 2. Within the automated procedure, maximum magnitudes for each zone are calculated using the cumulative strain energy method (Makropoulos and Burton, 1983), using seismic moment instead of strain energy. For an automated procedure, the maximum magnitude (M_3) is a good first order estimate of M_{MAX} that doesn't produce unphysical ($M_W > 10.0$) or logically inconsistent ($M_{MAX} < M_{MAX}^{OBS}$) upper limits on magnitude.

In addition to the use of existing source models, the synthetic catalogues must allow for the small but finite possibility that earthquakes greater than those found in the observed catalogue will occur. To do this a probabilistic switch between an assumed source model and regional seismicity is implemented. This is described by the following procedure:

1. Create N synthetic locations and magnitudes using the assigned source model.
2. Identify a randomly sampled (without replacement) subset of $0.05 \times N$ events
3. For all the events in the subset, sample magnitudes from a bounded Gutenberg and Richter relation defined by the regional catalogue parameters (M_{MIN} = 5.2, M_{MAX} = 8.6, $b = 1.20 \pm 0.026$ - for the declustered catalogue)

This method allows for a small but finite probability that earthquakes in the range $7.6 \leq M_W \leq 8.6$ (the regional maximum magnitude inferred by cumulative moment release in chapter 2) can be produced in the synthetic catalogue. It is a necessary step as it maintains the Poissonian behaviour, whilst allowing for the likelihood that the full rupture cycle for many faults and zones in the Aegean is not captured within the duration of the input catalogue. When creating a synthetic catalogue of many thousands of years in duration it is reasonable to expect that earthquakes with magnitudes approaching the maximum regional magnitude should be included, possibly several times. The approach described may be unrealistic in the sense that great earthquakes could feasibly be simulated around any observed earthquake location in the Aegean region. They are, however, more likely to be simulated in the areas of highest activity, which is around the Hellenic arc and Eurasian plate margin.

M_{MIN} is set to M_W 5.2: the completeness magnitude for the twentieth century Aegean catalogue. This means that the declustered catalogue for 1900 A.D. to 2005 A.D. can be input into the Monte Carlo program.

6.2.5 Perceptibility radius of earthquakes contributing to seismic hazard

It is possible, using the Monte Carlo procedure, to calculate the ground motion expected at a location from all the earthquakes in the synthetic catalogue. This is a computationally expensive procedure, as it will result in the prediction and analysis of many millions of ground motions that are insignificant, even for non-engineered structures. Furthermore, the predictive ground motion relations all have distance limits beyond which the ground motions are poorly modelled. This is especially true for relations lacking an intrinsic damping term. Of the attenuation relations used here none are considered to be accurate for source-site distances greater than 250 km.

Given these constraints, a "hazard radius" of 250 km is selected. This defines the distance beyond which an earthquake is considered to have a negligible impact on hazard at a site. This particular value is possibly a conservative estimate. It arises from the approximate maximum radius of macroseismic intensity (MMI) VI (the threshold of structural damage to class A buildings) predicted from the occurrence of a M_W 8.6 earthquake (the upper bound for the Aegean region), using the Papaioannou and Papazachos (2000) intensity attenuation relation for shallow earthquakes in the Aegean.

Earthquakes of large magnitude have caused significant damage over greater source-site distances in other parts of the world. The most obvious example is the 1985 Mexico City earthquake. Damage in these events is due to non-linear site amplification on unconsolidated sediments. There is little evidence in the historical record to suggest that damage from large events in the Aegean has been observed at such a considerable distance. Given that increasing the "hazard radius" would require a greater degree of extrapolation of attenuation relations beyond their derived maximum distance, the 250 km value is a reasonable compromise.

6.2.6 Mapping Seismic Hazard

The production of a seismic hazard map can introduce new sources of error, in addition to those inherent in the hazard calculation process. Initially the purpose of the hazard map must be determined. Hazard maps are usually unsuitable for the purposes of site-specific analysis. This is due to many of the simplifications in spatial characterisation of site and source. For the purposes of engineering design, a single value of PGA or spectral acceleration is insufficient. For the purposes of assessing the potential risk of an insurance portfolio, however, hazard maps are an invaluable resource. They can also be used to form a "seismic zonation" map, which provides a guide to implementation of building codes across a region.

The obvious question to ask is what parameter should a map use to characterise the hazard across a region. Traditionally, PGA has been the preferred strong motion parameter

for use in hazard mapping as it directly defines the inertial load force on a structure of known mass. Correlations of PGA with damage (typically MMI intensity) tend to be highly variable (Gutenberg and Richter, 1956; Krinitzsky and Chang, 1988; Trifunac and Brady, 1975) as PGA is not indicative of the frequency or the duration of strong shaking. This problem can be addressed by creating separate hazard maps for accelerations at specific spectral periods of interest (Petersen *et al.*, 2008). This can identify and distinguish between earthquake hazards from difference sources and illustrate the impact that larger, lower probability earthquakes have on a seismic hazard analysis.

PGA is not the only parameter that could be used in a seismic hazard map if the purpose of the map is to assess potential damage. Predictive attenuation relations are available for several other parameters of strong motion, such as those described in chapter 5. Several different options then become available. Perhaps the most obvious choice would be to use an intensity attenuation relation. This will calculate intensity directly based on the epicentral distance and the magnitude of the earthquake. The intensity attenuation relation used here is that of Papaioannou and Papazachos (2000), seen previously in Chapter 2:

$$I_{MM} = 1.43M_W - 3.59\log_{10}(R_{EPI} + 6) + 0.26 \pm 0.87 \tag{6.2}$$

where I_{MM} is the Modified Mercalli Intensity. Similar relations are proposed for intermediate-depth ($60 \leq h$ (km) ≤ 100) events:

$$I_{MM} = 1.69M_W - 3.34\log_{10}(R_{EPI} + 30) + 0.78 \tag{6.3}$$

and for deep events:

$$I_{MM} = 1.69M_W - 3.34\log_{10}(R_{EPI} + 30) - 0.69 \tag{6.4}$$

The use of intensity attenuation relations for the purposes of seismic hazard analysis can be insightful, but at the same time makes awkward assumptions about intensity data. Firstly macroseismic intensity is measured in an integer scale, which reflects an amount of damage across an undefined area. An intensity of VI assigned to a region will obviously mask local scale variations in strong shaking. Regression across a set of integer values requires determination of where in the integer scale each intensity level should sit (Musson, 1999a). Furthermore, different intensity scales have been used in the recording of historical earthquakes. The Modified Mercalli scale is used in the derivation of these attenuation relations. For more recent earthquakes in Europe the Medvedev-Sponheur-Karnik (MSK) and European Macroseismic Scale (Grünthal, 1998) are often used. Although correspondence between these scales is often treated as direct, this is not entirely correct. Different interpretations of damage, particularly in the intensity VI to VIII range, means that correspondence between scales is not direct. In lieu of correlation equations between scales, direct correspondence is assumed.

Several engineering based parameters could also be used in place of macroseismic intensity. Attenuation relations for several of these parameters (Travasarou *et al.*, 2003; Danciu and Tselentis, 2007) have been introduced in chapter 5. Many of these parameters are functions of the duration of strong shaking, as well as amplitude. Although not widely used by engineers, they often display better correlations to damage than PGA (Cabañas *et al.*, 1997). Empirical equations relating PGA, PGV, Arias Intensity and Cumulative Absolute Velocity to I_{MM} have recently been presented by Tselentis and Danciu (2008). These show considerable scatter and are suggested as valid only over a limited range of I_{MM} IV to VIII. Generally the uncertainty in these empirical relations translates to an uncertainty of approximately 1 level of Intensity.

The use of empirical equations relating ground motion scales to intensity is becoming more widespread. This is in no small part due to the adoption of such practice by the USGS in their Shakemap® software (Wald *et al.*, 2005): an automated rapid response program designed to give near real-time reports and forecasts of seismic strong ground motion following earthquakes. In addition to calculated PGA and PGV, "instrumental" intensity (intensity measured via empirical relation with a strong ground motion proxy) is also mapped using PGA as a proxy and the empirical relations of Wald *et al.* (1999) for California:

$$\begin{aligned} I_{MM} &= 3.66 \log_1 0\,(PGA) - 1.66 \quad \text{for} \quad V \leq I_{MM} \leq VIII \\ I_{MM} &= 2.20 \log_1 0\,(PGA) + 1.00 \quad \text{for} \quad I \leq I_{MM} < V \end{aligned} \quad \text{with} \quad \sigma_{MMI} = 1.08 \qquad (6.5)$$

This can be compared with the empirical relation of Tselentis and Danciu (2008), derived from Greek earthquakes (Figure 6.3):

$$I_{MM} = 3.563 \log_1 0\,(PGA) - 0.946 \quad \text{for} \quad IV \leq I_{MM} \leq VIII \quad \text{with} \quad \sigma_{MMI} = 0.734 \qquad (6.6)$$

In addition to the inherent variability (σ_{MMI}) in the PGA-I_{MM} empirical relations, which can be integrated into the Monte Carlo simulation, the treatment of intensity as a continuous scale is problematic. As an area of interest, hazard maps created using the Papaioannou and Papazachos (2000) MMI intensity attenuation relations will be compared with those produced using instrumental intensity for both the Wald *et al.* (1999) and Tselentis and Danciu (2008) relations.

Given the choice of various ground motion and damage parameters that have been considered, hazard maps will be produced using the following parameters: PGA (using various relations, see Chapter 5), Arias Intensity (Danciu and Tselentis, 2007)and Macroseismic Intensity (using Papaioannou and Papazachos (2000) and "instrumental" intensity). Other parameters such as PGV or CAV may also show significant correlation with damage and can readily be implemented in a Monte Carlo seismic hazard assessment, but are not considered here.

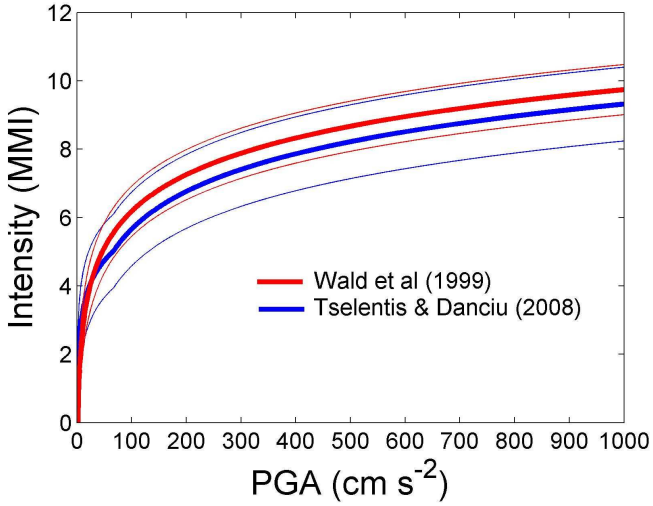

Figure 6.3: Comparison of the Intensity scaling relations of Wald *et al.* (1999) and Tselentis and Danciu (2008). Thin lines indicate $\pm 1\sigma$ relations

6.2.7 Resolution of calculated hazard points

As with many mapping applications the resolution of the mapping grid does have an influence on the final appearance of the map. A finer resolution may display greater detail and capture smaller variations in the seismic hazard. It does so, however, at a greater computational cost (Musson and Henni, 2001). In these maps hazard has been calculated on a grid with a spacing of $0.2^\circ \times 0.2^\circ$. This resolution represents a compromise between reasonable computational cost (between 2 and 4 hours to run each experiment on a desktop computer) and detail. For smooth contouring, hazard is interpolated onto 30 arc-second spaced grid using a bi-cubic spline algorithm.

6.3 "Single Model" Seismic Hazard Analysis

All these maps shown in this section assume a lower bound magnitude of M_W 5.2 and Poissonian seismicity. Seven models of seismicity (source and recurrence) are compared:

1. PP2000 with recurrence parameters calculated automatically from input catalogue.
2. PZ1990 with recurrence parameters calculated automatically from input catalogue.
3. WT2006 with recurrence parameters given in Chapter 3.
4. K27 (27-zone Rupture Cluster Analysis) model.
5. K29 (29-zone Rupture Cluster Analysis) model.
6. PP2000 with recurrence parameters given in Papaioannou and Papazachos (2000) (σ_b is assumed to be 0.01)

7. PZ1990 with recurrence parameters given in Papazachos (1990) (σ_b is assumed to be 0.01)

Uncertainty in b-value is incorporated by allowing b to vary according to a Gaussian distribution, with a standard deviation of σ_b. For earthquakes occurring outside the observed source zones, magnitudes are generated by random re-sampling with replacement from the subset of earthquakes found outside the source zones. No frequency magnitude distribution is assumed for these events, nor can magnitude exceed the maximum observed magnitude from the extra-zone earthquakes.

For each of the seven source models, six different hazard maps are produced using the following attenuation relations: Ambraseys *et al.* (1996), Skarlatoudis *et al.* (2003),Ambraseys *et al.* (2005a), Danciu and Tselentis (2007), Bommer *et al.* (2007) and Boore and Atkinson (2007). The hazard maps here assume ground motion a "rock" site (NEHRP Class B, V_{S30} = 760 $m\ s^{-1}$). Style-of-faulting is assumed to be strike-slip or normal/strike-slip, depending on how it is defined in each attenuation relation, except Boore and Atkinson (2007) where an "unknown" fault type is assumed.

Aleatory variability in the attenuation relation is taken into consideration using a Gaussian random number generator to add scatter to the logarithm of the estimated ground motion. The Monte Carlo method can allow for the variability to be unconstrained (i.e. with no limit on the number of standard deviations from the mean value). This can, however, result in unfeasibly high accelerations arising from small events. Here, the aleatory variability of ground motion is truncated at 3σ. This value is chosen on the guidance of several previous studies (Abrahamson, 2000; Bommer *et al.*, 2004; Abrahamson, 2006; Bommer and Abrahamson, 2006; Strasser *et al.*, 2008). The general consensus of these studies is that truncation at 1σ to 2σ underestimates the impact of aleatory variability within seismic hazard analysis. Conversely, even within extensive strong motion datasets, such as those used in the NGA project (Boore and Atkinson, 2007), the number of ground motion values in excess of 3σ from the 50^{th} percentile remain too small to constrain the tail ends of the distribution accurately. Analysis of the residuals demonstrates a deviation from the lognormal distribution for higher multiples of σ (Bommer *et al.*, 2004; Strasser *et al.*, 2008).

The distinction between inter- and intra-event variability is a much harder concept to integrate into a stochastic approach to seismic hazard analysis. Firstly, neither the Ambraseys *et al.* (1996) nor the Skarlatoudis *et al.* (2003) attenuation relations are generated using the random effects procedure. Consequently, inter- and intra-event variability cannot be characterised and comparison between the other relations is not necessarily valid. Instead for each strong motion calculation the total variability term σ_T is used to calculate aleatory uncertainty in strong ground motion modelling.

Hazard maps are shown for the PGA with a 10 % probability of being exceeded in 50 years (Figure 6.4). Hazard maps for the PGA with a 5 % and 2 % probability of being exceeded in 50 years can be found in Appendix B. The differences between these hazard maps clearly demonstrate the influence that both source model and attenuation relation have on the spatial distribution of seismic hazard. This illustrates the scale of the *epistemic uncertainty* associated with the hazard analyses in the Aegean region!

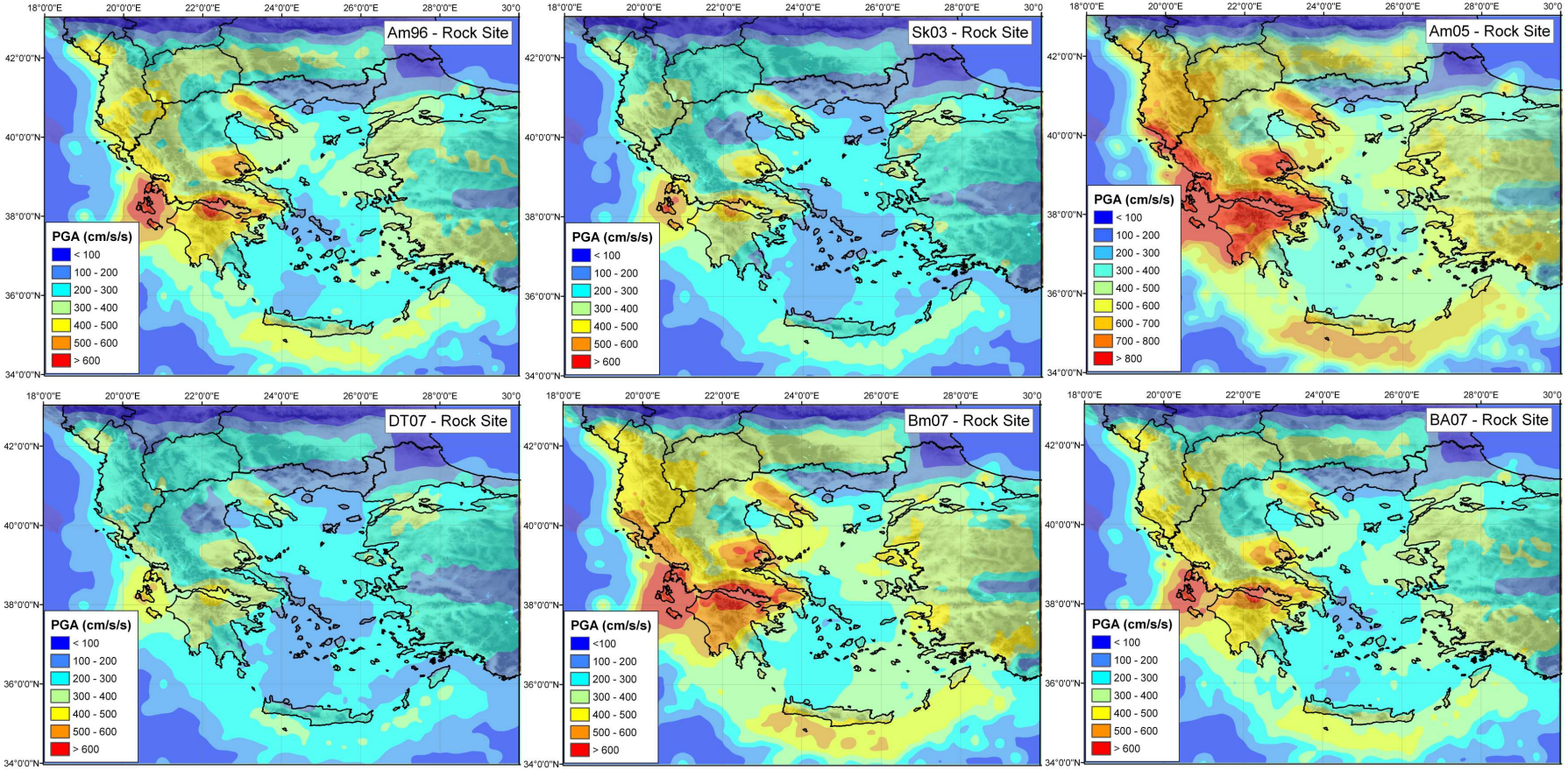

Figure 6.4: a) PGA with a 10 % probability of being exceeded in 50 years - PP2000(automatic) source model. Am96 (top left), Sk03 (top right), Am05 (centre left), DT07 (centre right), Bm07 (bottom left), BA07 (bottom right)

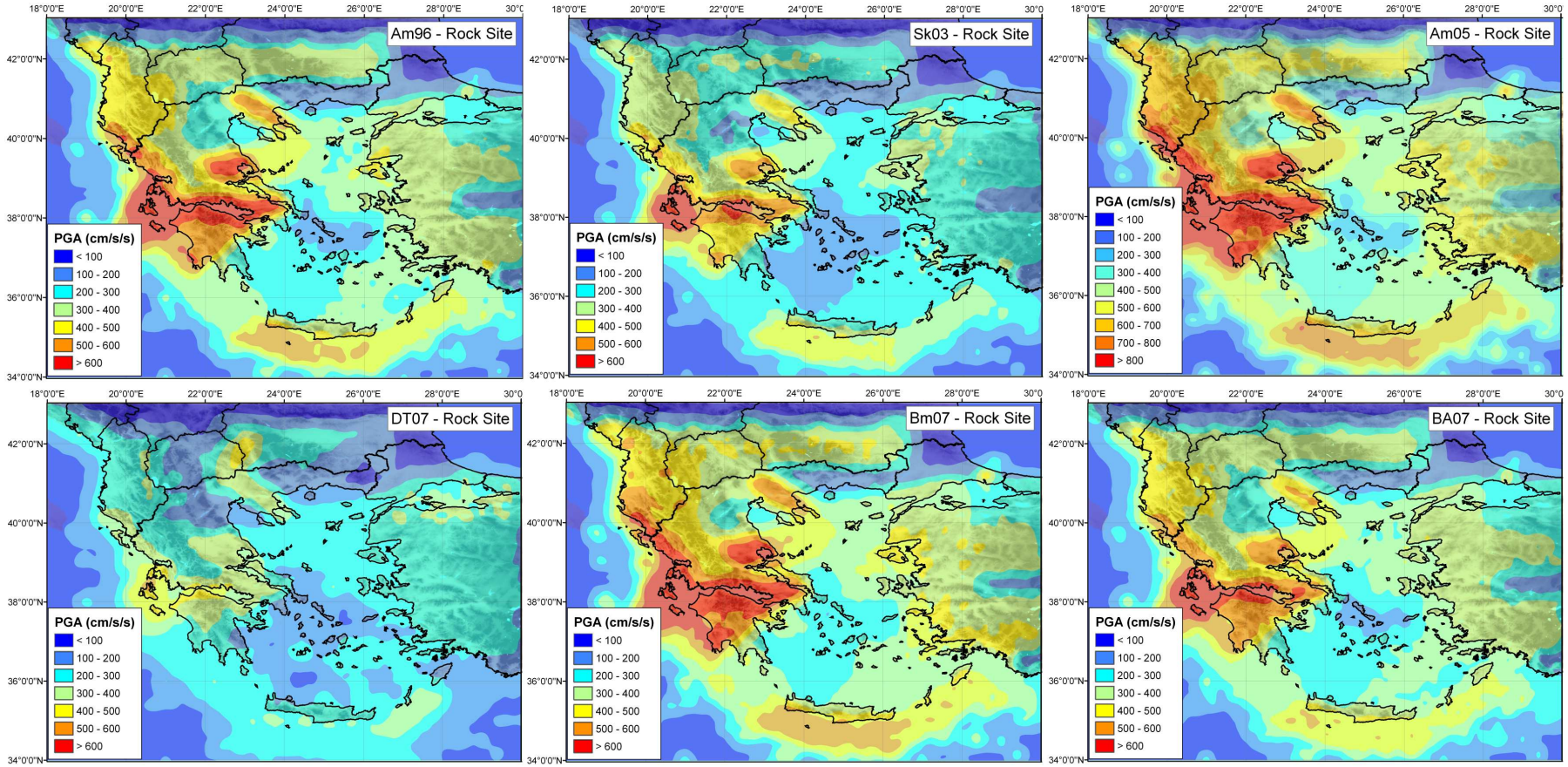

Figure 6.4: b) PGA with a 10 % probability of being exceeded in 50 years - PP2000(manual) source model. Am96 (top left), Sk03 (top right), Am05 (centre left), DT07 (centre right), Bm07 (bottom left), BA07 (bottom right)

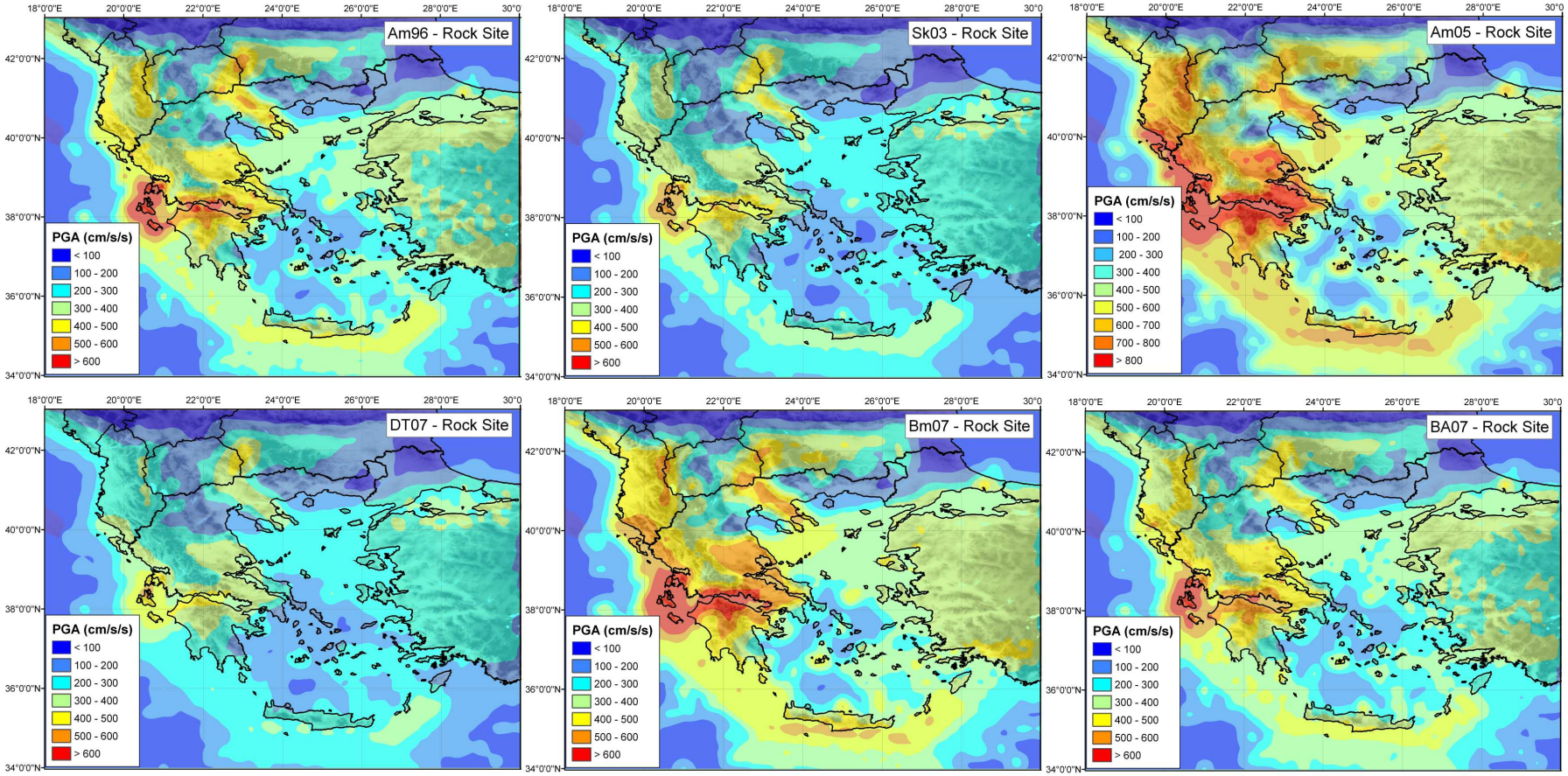

Figure 6.4: c) PGA with a 10 % probability of being exceeded in 50 years - PZ1990 (automatic) source model. Am96 (top left), Sk03 (top right), Am05 (centre left), DT07 (centre right), Bm07 (bottom left), BA07 (bottom right)

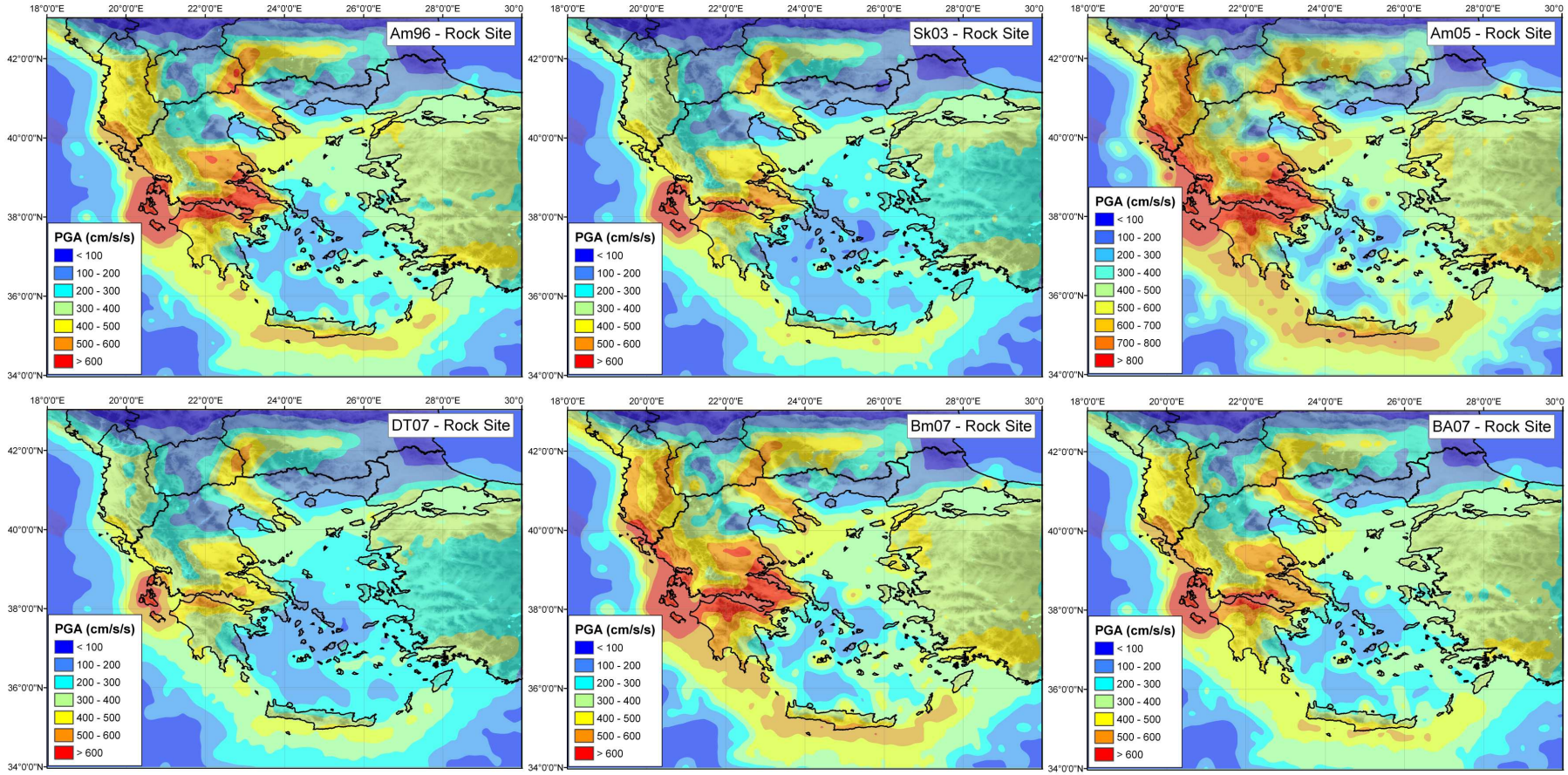

Figure 6.4: d) PGA with a 10 % probability of being exceeded in 50 years - PZ1990 (manual) source model. Am96 (top left), Sk03 (top right), Am05 (centre left), DT07 (centre right), Bm07 (bottom left), BA07 (bottom right)

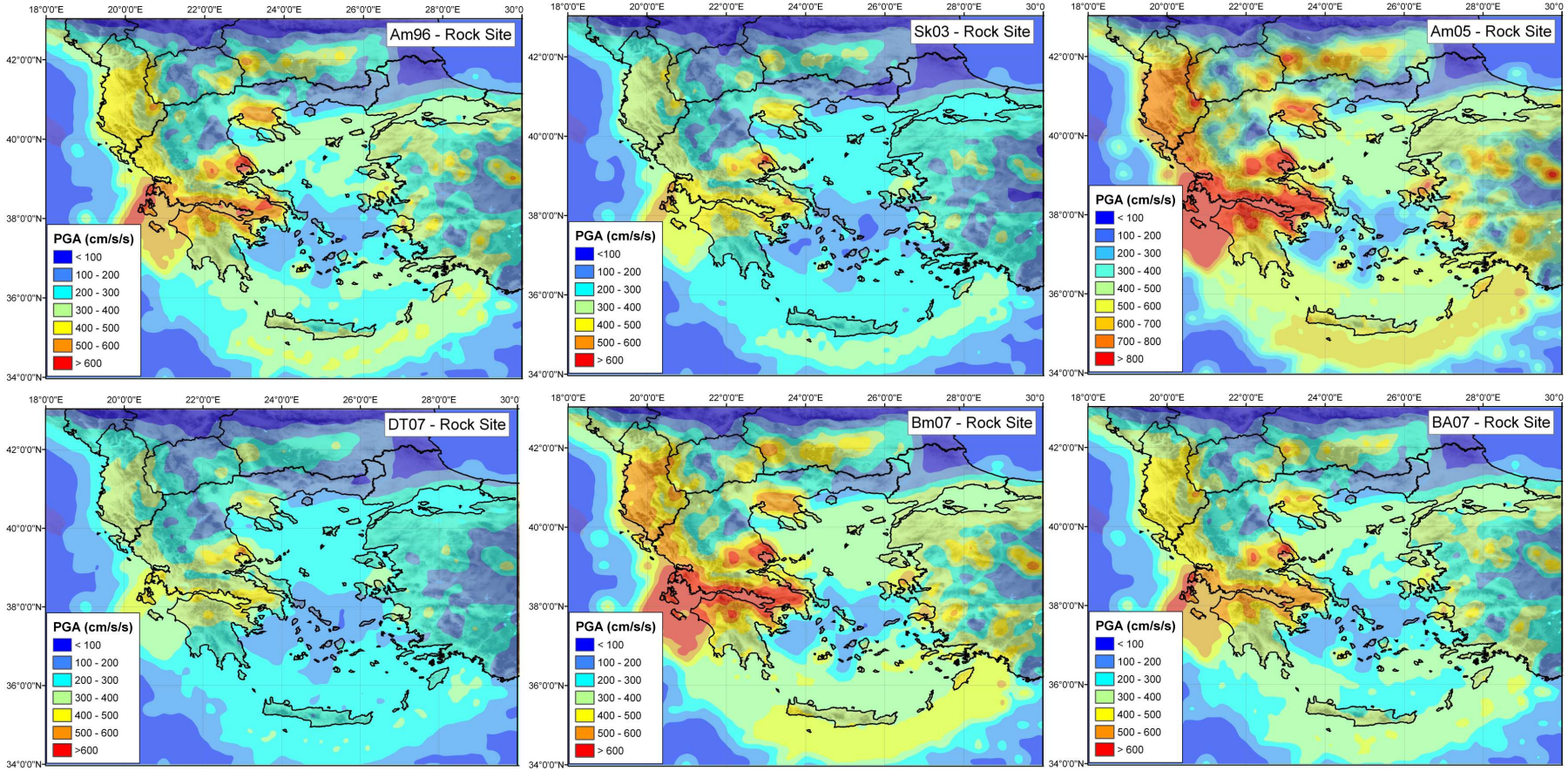

Figure 6.4: e) PGA with a 10 % probability of being exceeded in 50 years - WT2006 source model. Am96 (top left), Sk03 (top right), Am05 (centre left), DT07 (centre right), Bm07 (bottom left), BA07 (bottom right)

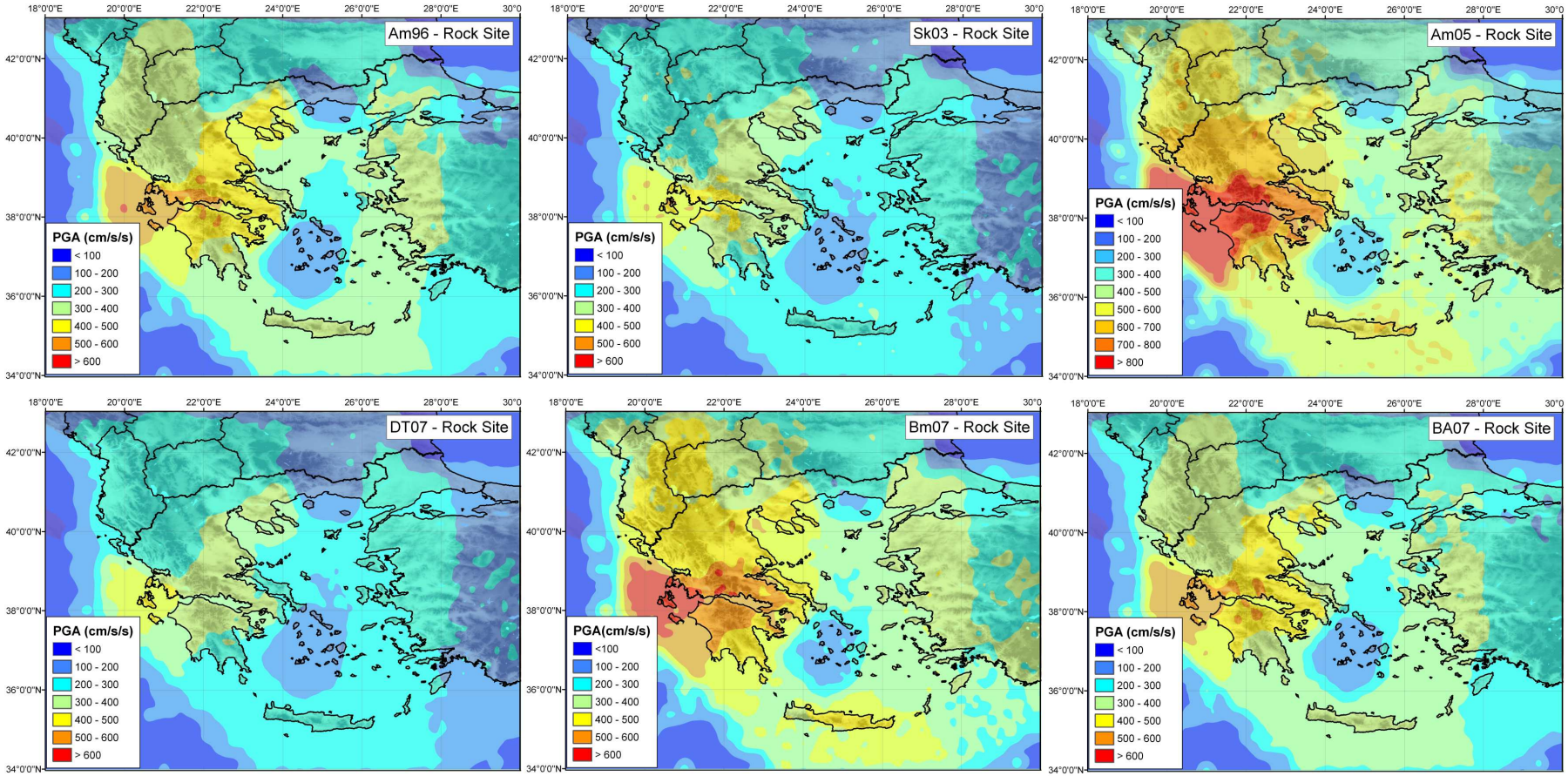

Figure 6.4: f) PGA with a 10 % probability of being exceeded in 50 years - K27 source model. Am96 (top left), Sk03 (top right), Am05 (centre left), DT07 (centre right), Bm07 (bottom left), BA07 (bottom right)

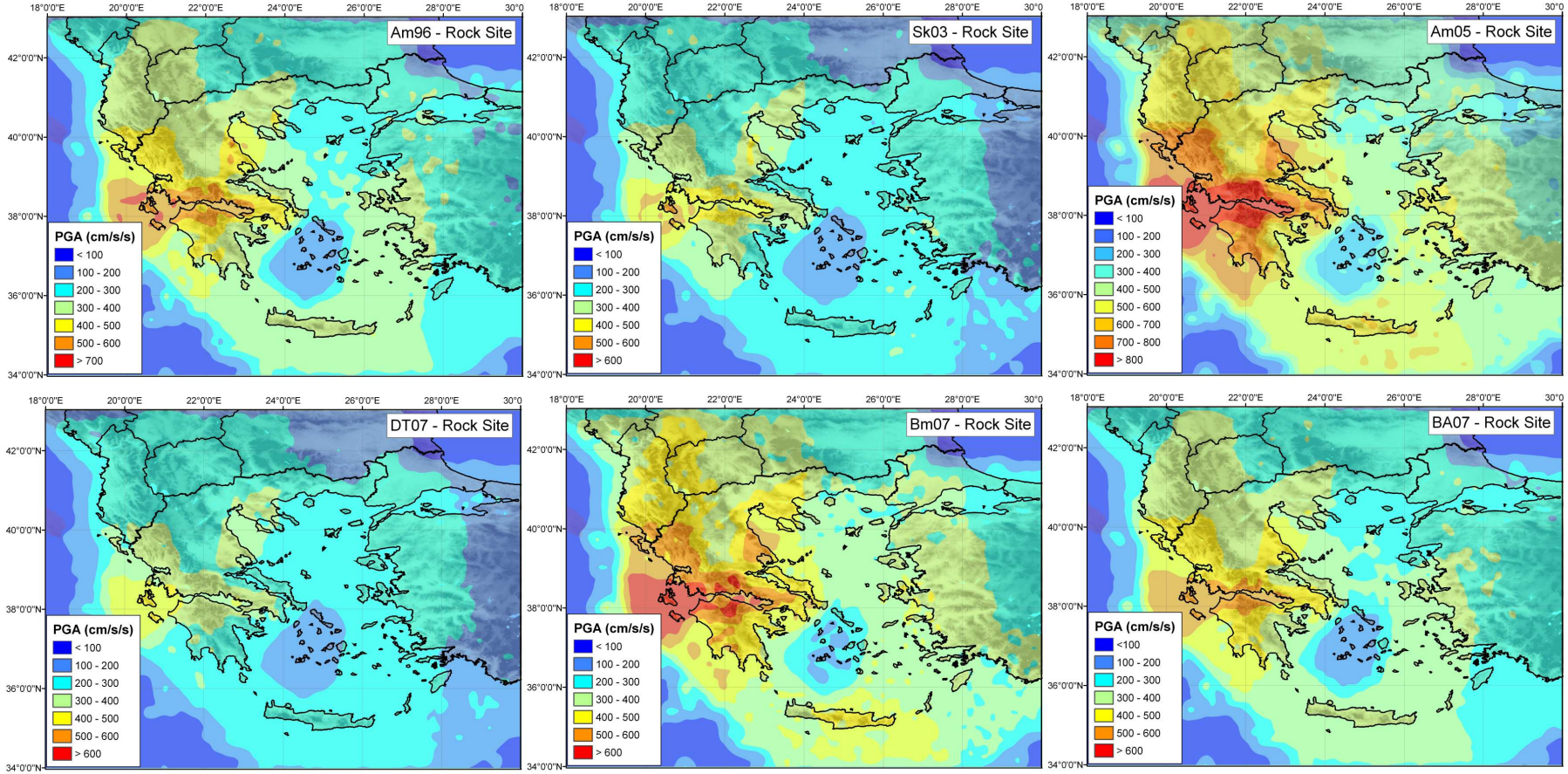

Figure 6.4: g) PGA with a 10 % probability of being exceeded in 50 years - K29 source model. Am96 (top left), Sk03 (top right), Am05 (centre left), DT07 (centre right), Bm07 (bottom left), BA07 (bottom right)

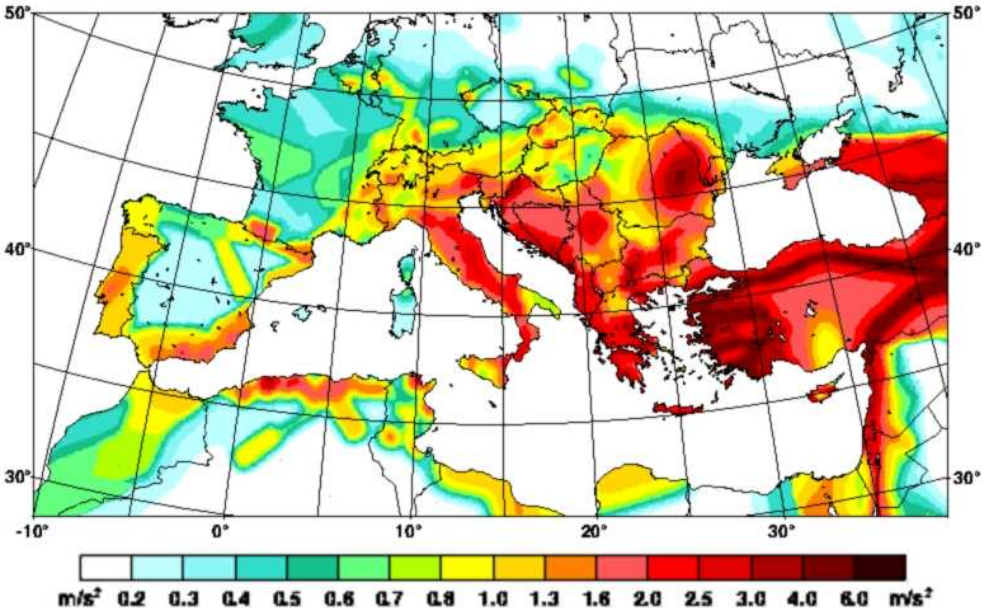

Figure 6.5: PGA with a 10 % probability of being exceeded in 50 years across Europe (GSHAP - Grünthal *et al.* (1999))

It is immediately clear that the choice of source model heavily influences the shape of the hazard maps. General consensus is that seismic hazard is higher in the Ionian Islands, the Gulf of Corinth and the Hellenic Arc. Similarly, hazard is lower in the South Aegean Sea and in F.Y.R.O.M. The highest ground accelerations in the Ionian Islands are mostly attributable to the high rate of seismicity, exacerbated by aleatory variability in the strong motion attenuation relation. These maps do not appear to contradict, strongly, other seismic hazard maps that have been developed for the Aegean region (e.g. GSHAP - Figure 6.5). The level of acceleration in these regions ($\approx$ 600 $cm\ s^{-2}$) is slightly larger than that predicted by GSHAP (Grünthal *et al.*, 1999) using the Ambraseys *et al.* (1996) relation, and slightly lower when using other relations.

The greatest contrast between these maps and those found in GSHAP, and similar analyses (Erdik *et al.*, 1999), is in Western Turkey. For most of the seismic hazard maps shown in Figure 6.4, the horizontal ground motion predicted in Western Turkey is lower than that predicted by Erdik *et al.* (1999), shown in Figure 6.6. Erdik *et al.* (1999) define narrow source zones around the most active faulting regions, and distributed seismicity elsewhere. This produces narrow bands of higher hazard around these zones. In contrast the zone models used in this analysis expand over larger areas, with a tendency to smooth the hazard over a wider area. Nevertheless, the PZ1990, WTH2006 and K27 models do forecast similar levels of ground motion across much of western Turkey (400 to 600 $cm\ s^{-2}$), with the exception of the eastern Sea of Marmara.

The hazard maps produced by the Monte Carlo method reflect the spatial variation in seismicity, but are heavily dependent on the source model. The WT2006 model is perhaps the most consistent with the expectations of the seismic hazard distribution given

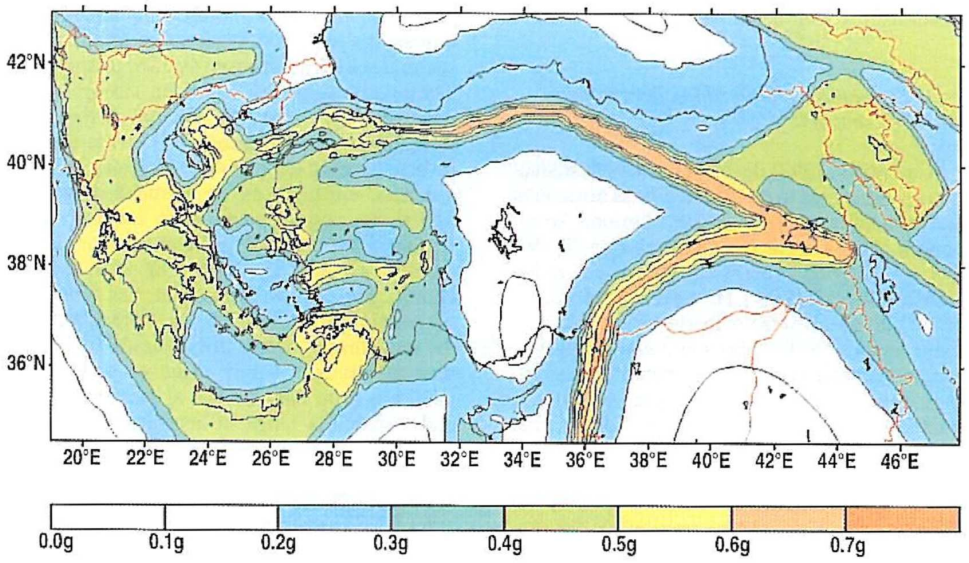

Figure 6.6: PGA with a 10 % probability of being exceeded in 50 years for Turkey and the surrounding regions Erdik *et al.* (1999)

the seismotectonic context. This is due to the delineation of large coherent zones around the North Anatolian fault system, rather than the segmentation of the fault system from east to west. In regions outside the uniform zones, the method of re-sampling and [slight] smoothing of observed seismicity to characterise the source produces more variable hazard maxima. Whether this is preferable to the zoning approach in these regions where faults are ill-constrained is debatable.

Given the disparity in the spatial distribution of hazard between different maps, how can one objectively assess which map is preferred? One approach may be to assess the similarity of each map to previous hazard analyses and form a judgement based on consistency. This could be considered a circular argument if previous analyses have not been validated in an objective manner. Similarly, if a common source model is used for this analysis and previous analyses (for example, PP2000 is common to this and GSHAP) then this model will be favourably biased over alternative models.

If previous hazard maps do not provide an objective means of validating source model, can observed seismicity and seismic hazard provide any better insight? This is discussed at length in Musson (2004), who uses synthetic catalogues for each zone to assess, via a χ^2 statistic, the distribution of b-values and mean magnitude compared against the observed values. This approach has been extended include consideration of strong motion in Chapter 4. Since hazard is a function of the spatial distribution of seismicity and magnitude distribution, it is harder to elucidate the best fitting models. This uncertainty is real, however, and since source models are commonly designed for the objective of hazard analysis, this comparative hazard approach is a useful extension. It also has an advantage as uniform zone source models can be compared with alternative source char-

acterisations (e.g. observed seismicity or smoothed seismicity).

As illustrated in Chapter 4, it is clear that source models that rely more on the observed distribution of seismicity will produce lower χ^2 values than uniform zones. If the χ^2 is calculated for the seven models considered (Figure 6.7), the reasons for this bias become obvious given that hybrid zone/zone-free models are being used. The WT2006 model has the lowest χ^2 value and the PP2000 the highest. The next lowest is PZ1990, whilst the K-means defined zones appear in the middle.

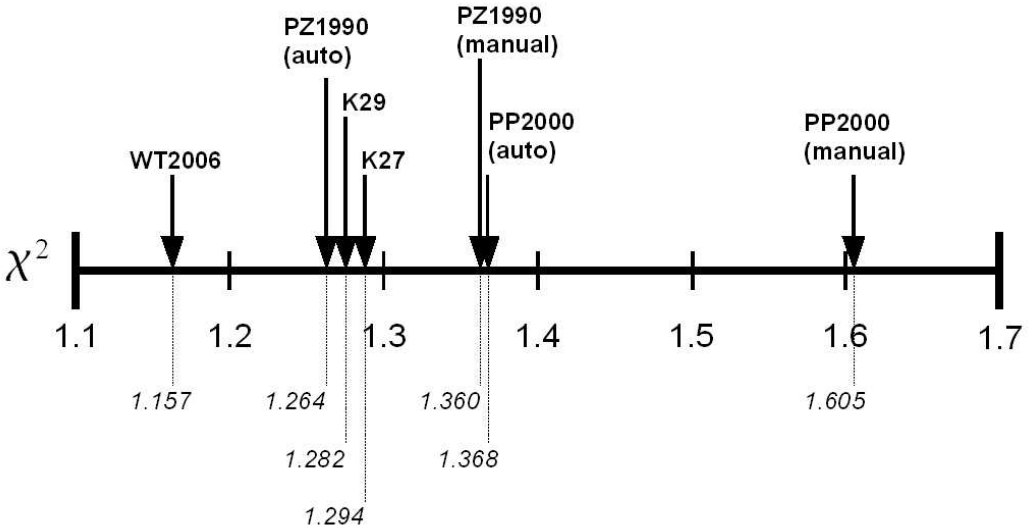

Figure 6.7: χ^2 values of fit to 100 years of observed seismicity for each source model

Both WT2006 and the PZ1990 models leave a substantial area of the Aegean region outside of a uniform zone. In the non-zoned areas it is the observed seismicity that characterises the source; hence synthetic seismicity will follow a similar distribution to the observed seismicity. The synthetic hazard calculated in these areas will therefore be similar to that observed. This clearly demonstrates why the hazard comparison method is not an objective method if comparing anything besides exclusively uniform models.

The absence of a truly objective means of validating source models is an important weakness in PSHA. These hazard maps clearly illustrate the influence different source models have on the hazard analysis, but ultimately the choice of model remains subject to user judgement. Unfortunately, a non-expert user of seismic hazard maps may not find this satisfactory. In the construction of national building codes it may not be sufficient to supply engineers and legislators with a plethora of maps, with considerable disparity, and suggest that the onus is on the user to make a judgement. It is obvious therefore that the characterisation of the seismic source is an important area of epistemic uncertainty.

Another interesting result in the hazard maps emerges when comparing the "automatic" and "manual" versions of the PP2000 and PZ1990 models. In the automatic model, a- and b- value are calculated from the observed seismicity in each zone using the max-

imum likelihood method (Aki, 1965). Maximum magnitudes are calculated using the cumulative moment method (Makropoulos and Burton, 1983). For zones with very few earthquakes, these parameters are estimated using inverse distance weighting from the nearest five zones. In the manual method, a, b and M_{MAX} for zones in the PP2000 and PZ1990 models are manually input from the parameters given in Papaioannou and Papazachos (2000) and Papazachos (1990) respectively. In both models the uncertainties on a- and b-value are not given, and are hence assumed to be 0.2 and 0.05 respectively. In the PP2000 model, the seismicity parameters given arise from the application of a kernel smoothing method (Papazachos, 1999a). This is designed to reduce the variability associated with short catalogues or transient features of seismicity, but it does so at the cost of removing the uncertainty estimate of b-value in these zones. The automatic and manual methods often produce different estimates of the recurrence parameters for each zone. However, when comparing the seismic hazard maps there is surprisingly little difference between these methods. This would suggest that the recurrence parameters have a less significant influence on hazard than either the source model or the ground motion model. The χ^2 comparison for these models in Figure 6.7 demonstrates that the "automatic" parameters provide a better fit to the observed data than the "manual" ones. Again, this is not surprising as the "automatic" parameters were derived directly from the observed data, whilst the "manual" ones were derived from a different data set and process. Since the smoothing process used by Papazachos (1999a) was implemented with the objective of reducing the impact of transient features of the catalogue on recurrence estimates, it is difficult to judge one set of parameters more preferable than any other for the same zonation scheme.

Whilst comparison of the source models and recurrence parameters provides interesting insight into the epistemic uncertainty prevalent in these hazard maps, the most obvious source of discrepancy in the levels of ground motion is in the ground motion prediction equation. Each of the attenuation relations shown produces a slightly different variation in hazard. It is immediately obvious that the three European relations (Am96, Bm05 and Bm07) predict higher levels of ground motion than their Greek (Sk03 and DT07) and NGA (BA07) counterparts. Of these it is Am05 that produces by far the highest levels of PGA across a wider area. In some places these values are approaching 1g. The lowest PGA's appear to be found using the DT07 relation.

The most obvious discrepancy between these PGA relations comes from the combination of horizontal components of strong ground motion. Am96, Am05 and Bm07 use the larger horizontal component of motion, whilst DT07 uses the geometric mean of the two components and BA07 the orientation-independent geometric mean. It is not clear how the horizontal components are combined in Sk03, but precedent by the same authors would suggest it is the geometric mean that is also used. This means that the accelerations forecast by the European relations should be systematically higher. If a correction factor of 1.14 is applied to PGA (Bommer *et al.*, 2005), the European relations can be adjusted down to the level of those using geometric mean horizontal component. This adjustment

is shown in Figure 6.8.

The scaling down of the European relations has an immediate impact that can be seen in the maps. Those maps created using Am96, Sk03, Bm07 and BA07 are now much closer in the level of ground motion exceeded than previously. In particular, hazard maps produced using the Bm07 and BA07 relations appear remarkably similar. Rescaling the relations to allow for comparison between larger horizontal component relations and geometric mean relations reduces the difference in hazard. There is little guidance as to which direction the scaling should be applied. Here, the decision has been made to scale down the European relations to bring them into line with the Greek relations. Whilst the definition of horizontal motion remains inconsistent, this is a dilemma that may have to be left to the demands of the user.

The issue of rescaling horizontal components of motion is also more complicated when considering hazard spectra. The scaling factor for median PGA of 1.14 used here (Bommer *et al.*, 2005) is not consistent across all spectral ordinates. It rises in a linear manner with the logarithm of period, reaching 1.27 at 4 s. The rescaling factor for standard deviation displays no obvious trend with period and is sufficiently close to unity (between 1 and 1.04) that no conversion is made here.

By making the adjustments to allow for better comparison between attenuation of larger component horizontal motion with geometric mean horizontal motion, some of the discrepancy between hazard maps has disappeared (Figure 6.8). There are still some remaining differences, the most obvious being the high accelerations found when the Ambraseys *et al.* (2005a) is used. Comparison of Am05 with the other attenuations relations was made in chapter 5. When considering the median level of ground motion there was good agreement between relations for larger earthquakes. Equally, the Am05, Bm07 and BA07 relations all show similar levels of median PGA for large earthquakes in the near field, and lower median PGA than Am96 and Sk03. Since Bm07 and Am05 are derived from a similar database of records then it cannot be argued that the high accelerations are a result of the data input.

Where there is some disparity between Am05 and the other relations is in the attenuation in the near field range of $1 \leq R_{JB}\ (km) \leq 15$. In this range the Am05 relation is remarkably level, predicting higher accelerations for smaller earthquakes than the other relations. This would suggest that with the Am05 relation, hazard may be controlled by smaller earthquakes in the near field range, an observation borne out by Musson (2006 - Personal communication).

Before jumping to this conclusion consideration needs to be given to the aleatory uncertainty term. Both Am05 and Bm07 have heteroscedastic aleatory variability. If this were the sole influence then similar hazard levels from these two relations would not be unexpected. Figure 6.9 also shows that although aleatory variability is greater for Am05 for

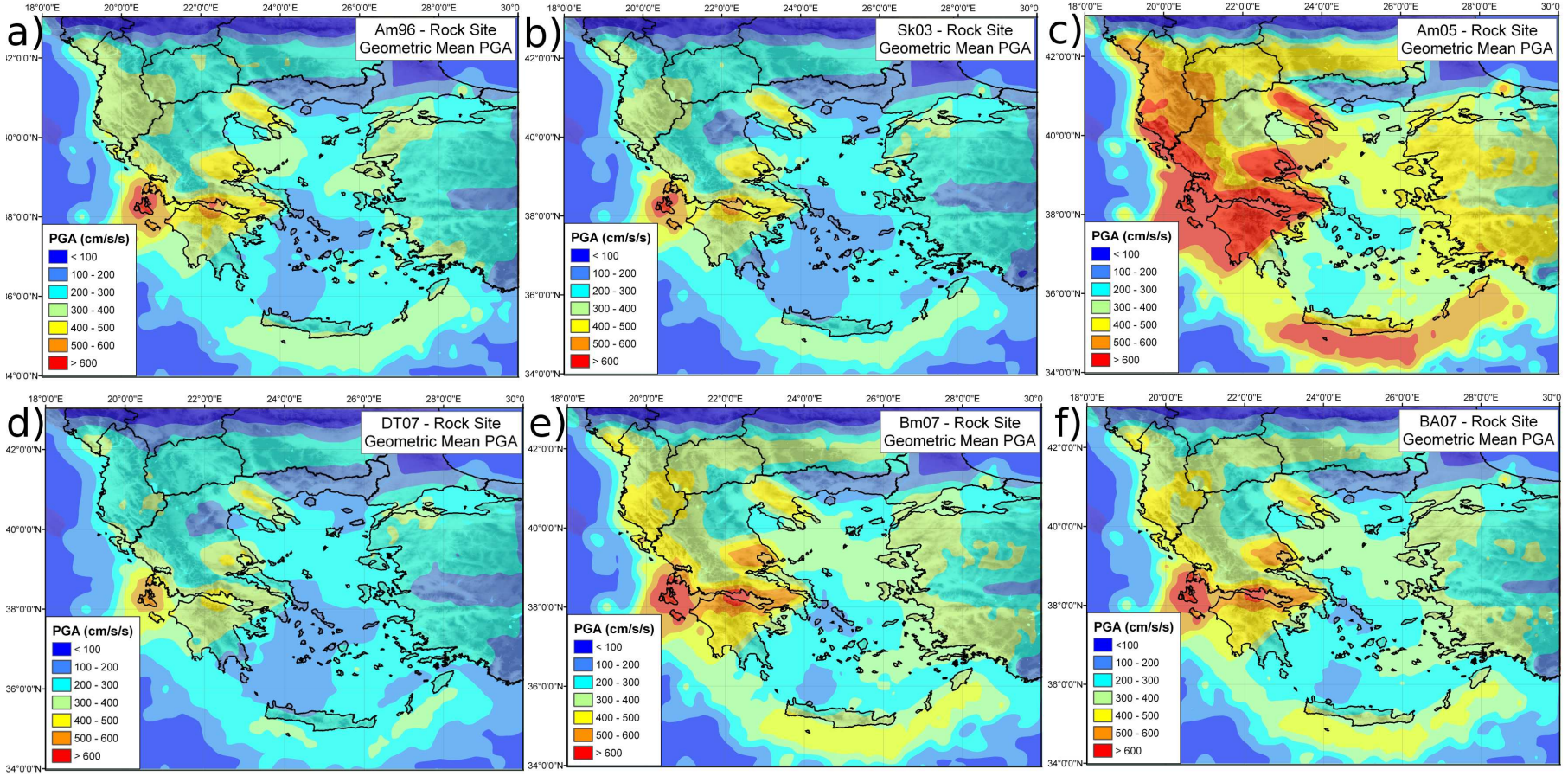

Figure 6.8: As Figure 6.4a) with larger horizontal component PGA relations (Am96, Am05 and Bm07) scaled down for comparison with geometric mean horizontal PGA.

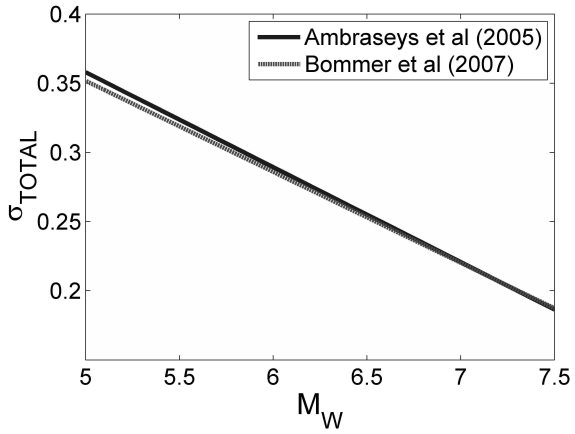

Figure 6.9: Comparison of heteroscedastic aleatory variability for the Am05 and Bm07 relations.

smaller earthquakes, the differences are not sufficient to explain why the Am05 hazard map is so much higher than the Bm07 map. Furthermore, even the high sigma values associated with smaller earthquakes in the Am05 relation ($\sigma \approx 0.35$) are much smaller than the fixed sigma value found in BA07 ($\sigma = 0.502$). This suggests that higher aleatory variability alone doesn't translate into larger hazard at this return period, though it may do so for longer return periods. The higher levels of PGA with a 10 % probability of being exceeded in 50 years, using the Am05 relation, can be attributed mostly to the high accelerations in the near-field, and to a lesser extent the magnitude-dependent aleatory uncertainty.

The explanation for the high PGA values in the Am05 hazard map also helps account for the low PGA values in the DT07 map. The relation with lower PGA predicted for smaller earthquakes, and a smaller aleatory uncertainty term, would produce lower values of hazard overall. This is the case with the DT07 relation, which predicts substantially lower accelerations for small to moderate events than its European counterparts. The aleatory uncertainty term is also fixed at 0.291; an intermediate value in the range found in the Am05 and Bm07 relations.

6.4 Hazard Maps for Alternative Ground Motion Parameters

So far, only PGA has been considered as the strong motion parameter. Alternative strong motion parameters, macroseismic intensity in particular, may have interesting implications. Comparison of intensity maps using different methods and proxies can be seen in Figure 6.10. Intensity maps are derived from exclusively shallow earthquakes and are constructed in six different ways:

1. Using the intensity attenuation relation of Papaioannou and Papazachos (2000).

2. Using the Wald *et al.* (1999) empirical relation between PGA (determined directly using BA07) and I_{MM} (Equation 6.4). EMS/MSK intensity is assumed equivalent to I_{MM} for this purpose.

3. Using the Tselentis and Danciu (2008) [TD08] "simple" empirical relation between PGA and I_{MM} (Equation 6.5)

4. Using the TD08 "site adjusted" empirical relation between PGA and I_{MM}

$$I_{MM} = 2.355 + 1.384 \log_{10}(PGA) + 0.297 M_W - 0.832 \log_{10}(R_{EPI}) - 1.088 S + 0.666\sigma \tag{6.7}$$

   where S = 0 on "rock" (NEHRP Class B) sites, S = 1 on "stiff soil" and S = 2 on "soft soil".

5. Using the TD08 "simple" empirical relation between Arias Intensity (I_a) (calculated directly using DT07) and I_{MM}:

$$I_{MM} = 4.395 + 2.040 \log_{10}(I_a) + 1.278\sigma \tag{6.8}$$

6. Using the TD08 "site adjusted" empirical relation between Arias Intensity and I_{MM}

$$MMI = 5.919 + 0.844 \log_{10}(I_a) - 0.997 \log_{10}(R_{EPI}) - 0.105 S + 0.649\sigma \tag{6.9}$$

The use of a "Greek specific" attenuation relation in method 1 allows for comparison of I_{MM} hazard maps with hazard maps created using the SGM proxies. The production of I_{MM} maps using PGA and I_a proxies is also illustrate how a Monte Carlo technique can be made compatible with the ShakeMap methodology. This is done directly using the Wald *et al.* (1999) empirical relations, which are widely used in ShakeMap. The use of these proxies adds an additional element of uncertainty owing to the scatter of the residuals in the PGA-I_{MM} relations. This additional uncertainty has been assimilated into the calculations using Gaussian scatter in the Monte Carlo simulations. Only the PP2000 source model is used in these maps.

These intensity hazard maps show the considerable variation when the different methods are used. The use of the Papaioannou and Papazachos (2000) intensity attenuation relation predicts that for most of the continental Aegean the I_{MM} with a 10 % probability of being exceeded in 50 years is VII. Only the Ionian-Corinth region is higher at intensity VIII. Using PGA and the Wald *et al.* (1999) relation these same regions are now a level higher (IX in the Ionian-Corinth region, VIII across much of the Aegean). When using Arias Intensity as a proxy for the calculation of I_{MM}, the resulting macroseismic intensities are even higher (I_{MM} = X in central Greece). It is also interesting to note the impact that the site condition factor introduced by Tselentis and Danciu (2008) has on the I_{MM}

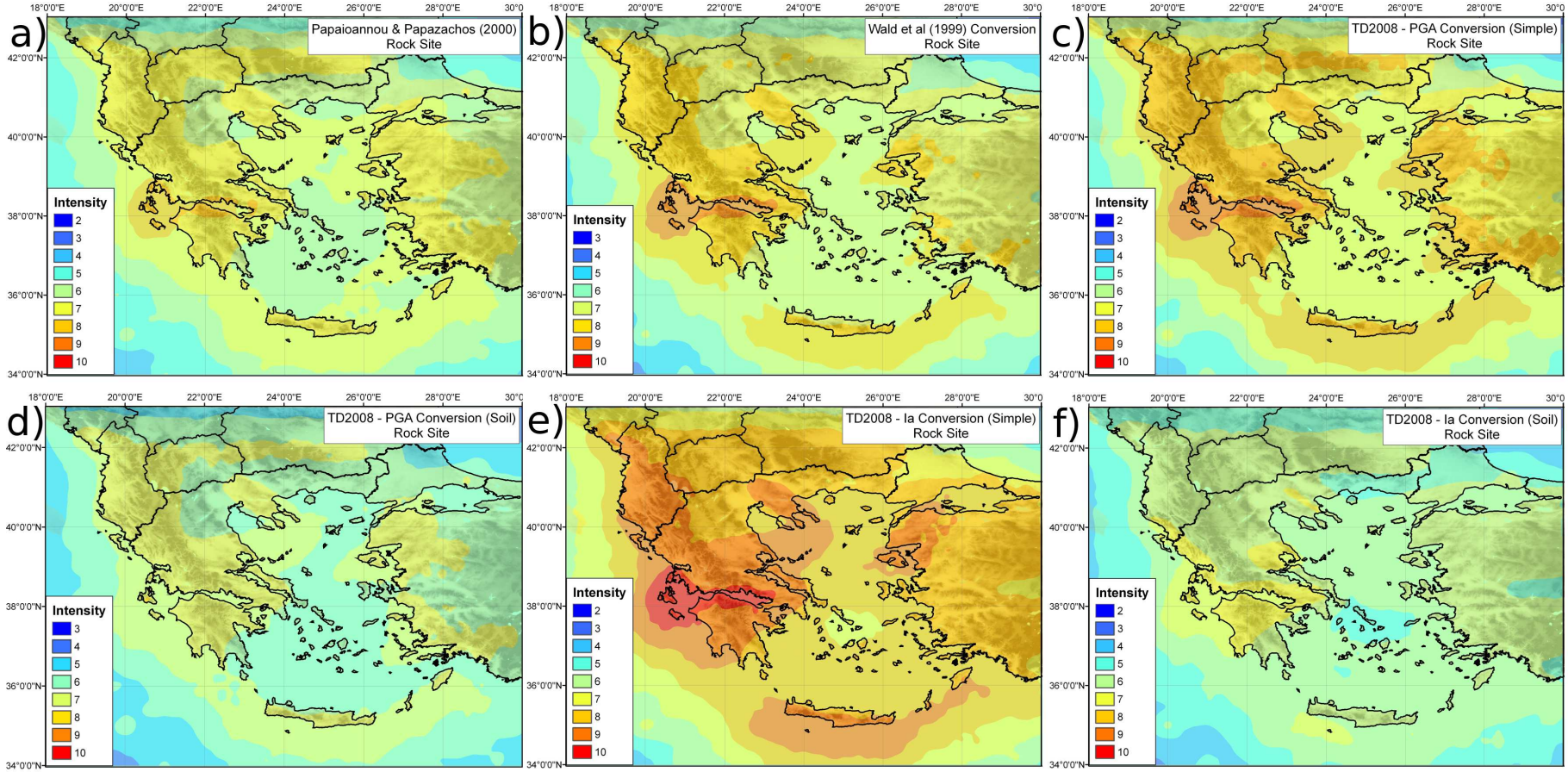

Figure 6.10: Modified Mercalli intensity with a 10 % probability of being exceeded in 50 years using the PP2000 source model. a) Papaioannou and Papazachos (2000) (top left), b) I_{MM} via the Wald *et al.* (1999) empirical relation (top right), c) I_{MM} using the Tselentis and Danciu (2008) relation [TD08] d) PGA and I_{MM} empirical relation (centre left), I_{MM} using TD08 PGA and I_{MM} empirical relation, with adjustment for site characteristic, e) I_{MM} using the TD08 Arias Intensity to I_{MM} empirical relation (bottom left), f) I_{MM} using the TD08 Arias Intensity to I_{MM} empirical relation, with adjustment for site characteristic (bottom right)

hazard maps. These reduce the intensity by as much as 2 intensity levels compared to the "simple" relation.

Although these maps show a disparity, the agreement is not altogether unreasonable when the variability in the PGA-I_{MM} conversion relations is taken into account. The sigma values in these relations are such that the 5 % to 95 % confidence interval translates into a difference of almost three intensity levels. In the $VI \leq I_{MM} \leq IX$ range, this can mean the difference between minor damage in isolated regions and extensive damage over a wide area. A fairer assessment may be reached when taking into account spatial variation in site condition, which will be considered in the next chapter.

6.5 Application of the Monte Carlo Technique to Site-Specific Seismic Hazard Analysis

6.5.1 Site-Specific Analysis

Seismic hazard maps provide important information regarding the spatial variation in seismic hazard and the likely areas to be affected by strong shaking. There is, however, a limit to the amount of information than can be assimilated into a map, or even a suite of maps such as those seen here. Visualisation of spatial variation in hazard has additional uncertainties arising from interpolation and contouring. Site-specific seismic hazard analyses require a different procedure from a mapping application, even if the underlying theory remains the same. It must be emphasised that what is being presented here is the application of a Monte Carlo based probabilistic seismic hazard analysis to a reference location point, with an accompanying "site classification" value. This is not identical to the sort of site-specific hazard investigation that is demanded for seismically sensitive engineered structures. Such an analysis would require a greater degree of investigation into both the site itself and the contributing seismic sources, something which is beyond the level of detail considered here. Nevertheless, the Monte Carlo method can produce much of the information needed as input for more detailed analyses, such as hazard curves and design spectra.

6.5.2 Seismic Hazard Curves

For the eight sites listed in Table 6.1 hazard curves are shown in Figure 6.11. Separate curves are shown for five attenuation relations (Am96, Am05, DT07, Bm07, BA07), assuming the same source model. The values of PGA with a 10 %, 5 % and 2 % probability of being exceeded at each site are given in Table 6.2. The curves show the variation in annual probability of ground motion being exceeded, with PGA. This is derived from the same assumption of Poissonian distributed ground motion as that seen in Equation 6.11, this time with T = 1.

The faulting parameter in the attenuation relation is assumed to be normal or normal/strike-slip, depending on how it is defined in the attenuation relation. A comparison of the hazard curves for all 8 sites, assuming the PP2000 source model, can also be seen in Figure 6.12. A comparison of the hazard curves for different source models will be discussed later on in this section.

Table 6.2: PGA ($cm\ s^{-2}$) with a 10 %, 5 % and 2 % probability of being exceeded in 50 years (P.B.E. 50), for 8 major cities in the Aegean region

P.B.E. 50	PP2000				
10 %	Am96	Am05	DT07	Bm07	BA07
Argostoli	946.93	1202.31	593.58	790.44	774.81
Athens	464.29	642.04	294.29	449.60	381.43
Heraklion	368.76	514.42	230.06	361.71	318.19
Istanbul	259.88	299.84	208.61	241.00	242.13
Izmir	365.21	606.16	247.19	390.06	379.80
Rhodos	344.66	500.85	214.70	340.82	307.97
Thessaloniki	469.76	762.21	347.23	484.50	482.79
Tirane	477.09	682.18	296.67	458.88	422.70

5 %	Am96	Am05	DT07	Bm07	BA07
Argostoli	1131.80	1469.66	779.02	946.84	912.30
Athens	614.99	856.71	381.98	564.95	476.91
Heraklion	493.73	679.20	309.80	471.13	416.20
Istanbul	363.86	425.99	269.05	319.98	314.53
Izmir	480.07	814.29	320.79	516.66	493.57
Rhodos	454.14	670.52	281.80	445.15	394.98
Thessaloniki	622.52	1015.66	474.87	655.22	574.75
Tirane	626.64	881.21	364.97	585.60	518.58

2 %	Am96	Am05	DT07	Bm07	BA07
Argostoli	1468.57	1871.24	1001.59	1179.98	1123.30
Athens	822.38	1168.85	534.41	747.95	634.63
Heraklion	744.08	945.32	446.45	641.83	602.77
Istanbul	577.51	610.60	419.79	450.72	456.35
Izmir	634.19	1139.15	432.66	695.82	593.73
Rhodos	629.97	947.42	402.74	617.26	514.27
Thessaloniki	932.68	1380.80	702.64	884.90	741.42
Tirane	873.66	1119.72	494.86	773.40	705.32

P.B.E. 50	PZ1990				
10 %	Am96	Am05	DT07	Bm07	BA07
Argostoli	922.17	1253.70	595.18	816.98	765.25
Athens	275.16	385.98	207.84	283.92	248.15
Heraklion	447.88	665.82	277.58	438.51	391.49
Istanbul	359.64	417.44	273.77	319.45	315.52
Izmir	416.25	608.48	301.66	415.97	415.56
Rhodos	327.74	477.28	198.85	329.02	289.09
Thessaloniki	501.71	821.60	359.91	522.94	489.01
Tirane	511.74	720.77	310.87	489.66	446.28

5 %	Am96	Am05	DT07	Bm07	BA07
Argostoli	1170.01	1509.88	776.07	977.29	922.80
Athens	342.49	497.25	257.59	345.29	302.09
Heraklion	583.31	849.09	349.79	556.84	503.07
Istanbul	508.03	570.26	370.94	424.00	390.49
Izmir	581.92	851.49	402.34	553.85	528.22
Rhodos	439.90	683.69	252.50	431.17	381.12
Thessaloniki	655.09	1017.34	454.26	665.99	577.22
Tirane	657.52	899.04	392.32	610.06	559.57

2 %	Am96	Am05	DT07	Bm07	BA07
Argostoli	1590.52	1955.74	1066.60	1235.93	1193.06
Athens	446.59	690.17	360.08	460.02	369.62
Heraklion	772.08	1153.19	514.45	726.10	663.31
Istanbul	698.75	780.07	530.04	565.49	534.17
Izmir	836.02	1139.74	618.08	772.62	732.27
Rhodos	564.69	928.93	367.68	576.94	527.14
Thessaloniki	831.13	1426.53	622.76	804.59	724.89
Tirane	849.94	1213.02	502.96	784.04	703.38

P.B.E. 50	WT2006				
10 %	Am96	Am05	DT07	Bm07	BA07
Argostoli	843.31	1132.48	512.89	740.56	723.00
Athens	606.05	830.73	372.24	550.47	483.21
Heraklion	351.12	523.04	225.42	355.98	320.65
Istanbul	374.74	456.37	256.39	328.60	325.08
Izmir	553.24	905.21	356.63	547.42	527.06
Rhodos	520.69	703.42	324.59	472.02	444.28
Thessaloniki	680.40	987.87	456.68	653.86	584.22
Tirane	556.43	840.68	354.97	553.52	492.98

5 %	Am96	Am05	DT07	Bm07	BA07
Argostoli	1057.79	1421.48	655.58	909.17	887.35
Athens	767.54	1049.01	487.78	686.99	607.88
Heraklion	473.52	702.82	290.71	464.56	404.19
Istanbul	530.00	602.62	356.45	438.88	416.78
Izmir	737.91	1187.14	470.20	713.89	676.37
Rhodos	696.62	931.97	430.43	624.50	551.13
Thessaloniki	866.79	1244.52	574.10	801.04	696.15
Tirane	723.48	1072.11	451.42	679.73	626.96

2 %	Am96	Am05	DT07	Bm07	BA07
Argostoli	1414.05	1836.07	983.10	1148.04	1126.71
Athens	1011.35	1397.48	682.69	932.17	791.42
Heraklion	676.21	923.36	399.84	599.94	544.89
Istanbul	790.59	902.69	519.70	621.05	604.19
Izmir	979.32	1581.79	702.83	916.16	878.11
Rhodos	899.23	1289.96	622.87	856.02	775.02
Thessaloniki	1130.55	1618.47	822.23	1048.02	869.02
Tirane	970.43	1448.59	609.87	893.45	801.65

P.B.E. 50	K27				
10 %	Am96	Am05	DT07	Bm07	BA07
Argostoli	753.61	1014.14	472.74	654.73	634.29
Athens	550.08	814.45	330.56	527.59	456.52
Heraklion	467.05	628.10	274.76	438.43	386.97
Istanbul	264.81	314.28	196.45	237.14	242.70
Izmir	431.92	687.51	307.57	448.04	432.30
Rhodos	333.32	484.50	215.36	328.69	295.82
Thessaloniki	542.48	763.06	382.08	517.19	486.08
Tirane	453.50	604.86	270.39	415.47	414.20

5 %	Am96	Am05	DT07	Bm07	BA07
Argostoli	941.32	1226.24	648.81	792.22	782.98
Athens	687.69	1040.27	420.95	655.69	561.66
Heraklion	605.73	842.89	372.53	552.44	511.94
Istanbul	382.08	451.54	272.64	339.18	322.98
Izmir	576.01	868.09	414.11	567.13	567.15
Rhodos	459.49	639.87	292.74	430.66	385.10
Thessaloniki	713.06	981.34	536.53	648.60	603.76
Tirane	598.63	797.22	376.60	533.90	512.22

2 %	Am96	Am05	DT07	Bm07	BA07
Argostoli	1259.59	1518.20	841.45	984.33	986.66
Athens	902.86	1335.88	577.19	865.16	743.82
Heraklion	878.40	1122.26	487.82	732.88	641.90
Istanbul	591.07	653.03	416.44	495.82	450.59
Izmir	843.48	1255.14	598.97	816.91	737.21
Rhodos	668.37	931.23	389.42	588.71	537.02
Thessaloniki	1006.45	1272.40	777.08	838.74	727.71
Tirane	827.66	1115.68	494.85	716.19	696.31

P.B.E. 50	K29				
10 %	Am96	Am05	DT07	Bm07	BA07
Argostoli	743.35	1039.53	464.89	675.88	638.01
Athens	556.94	811.28	327.67	528.12	455.38
Heraklion	453.10	649.26	275.38	432.35	393.77
Istanbul	270.37	322.22	190.94	237.82	245.35
Izmir	481.00	741.71	340.37	489.59	466.41
Rhodos	332.39	505.64	210.50	334.63	294.00
Thessaloniki	518.23	770.14	367.10	523.42	483.98
Tirane	485.45	633.98	293.10	437.20	425.06

5 %	Am96	Am05	DT07	Bm07	BA07
Argostoli	936.60	1267.16	616.31	803.14	779.56
Athens	720.86	1004.18	438.22	653.43	555.62
Heraklion	580.28	864.01	359.87	548.34	495.41
Istanbul	370.56	464.70	271.74	339.93	319.38
Izmir	672.86	954.36	436.00	615.28	587.78
Rhodos	420.34	666.78	289.53	426.61	369.02
Thessaloniki	673.00	1062.38	477.37	666.94	590.35
Tirane	624.74	790.68	382.38	549.58	536.21

2 %	Am96	Am05	DT07	Bm07	BA07
Argostoli	1205.09	1529.36	849.66	990.35	993.55
Athens	941.88	1393.80	620.25	873.98	725.34
Heraklion	815.12	1125.28	467.96	755.58	645.04
Istanbul	600.98	722.60	423.12	477.58	444.74
Izmir	911.14	1289.82	627.38	814.89	772.46
Rhodos	614.57	911.39	403.12	602.32	499.24
Thessaloniki	895.53	1494.75	647.81	894.15	755.82
Tirane	829.03	1012.60	509.10	695.32	659.98

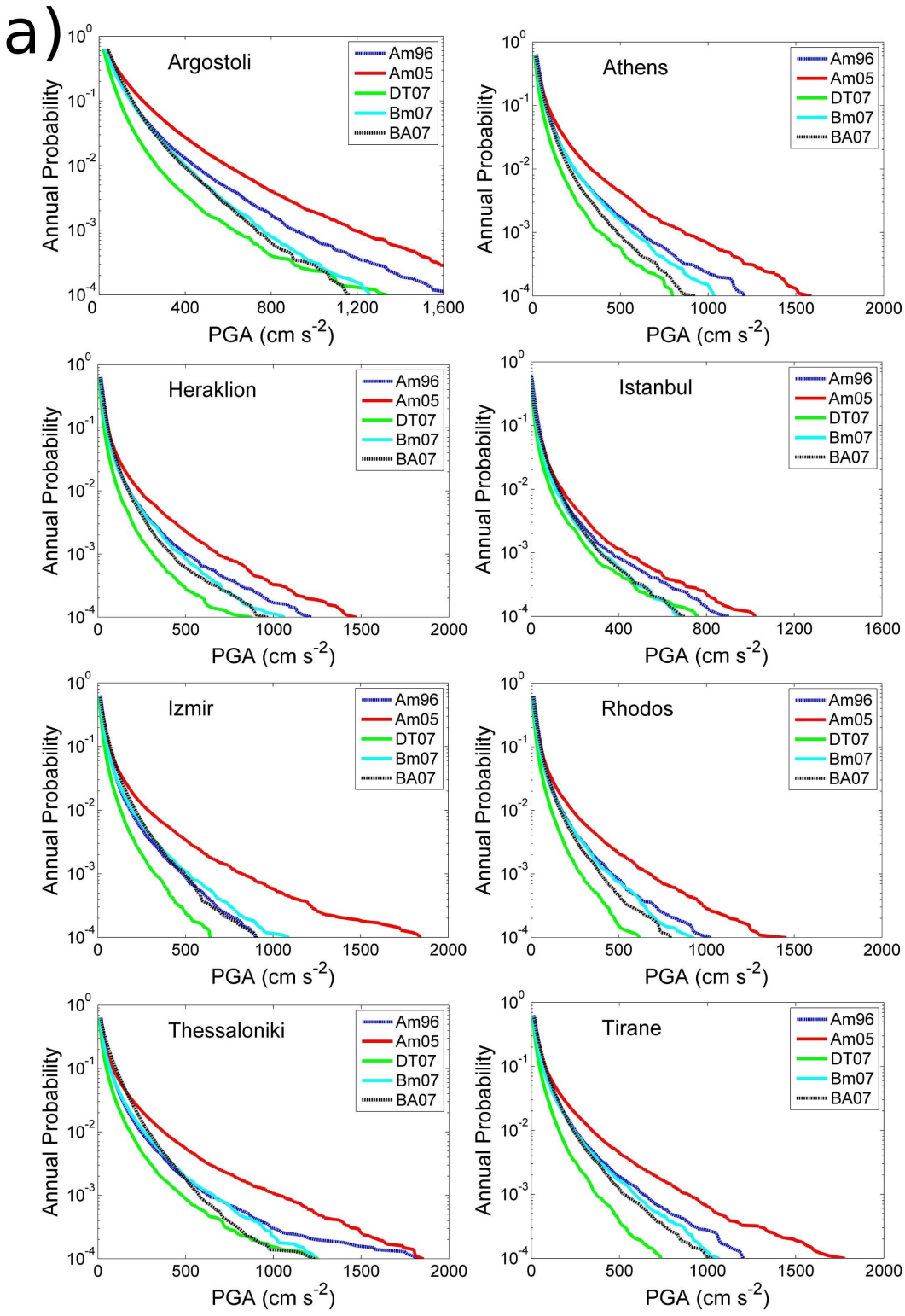

Figure 6.11: a) Seismic Hazard Curves - PP2000 Source Model

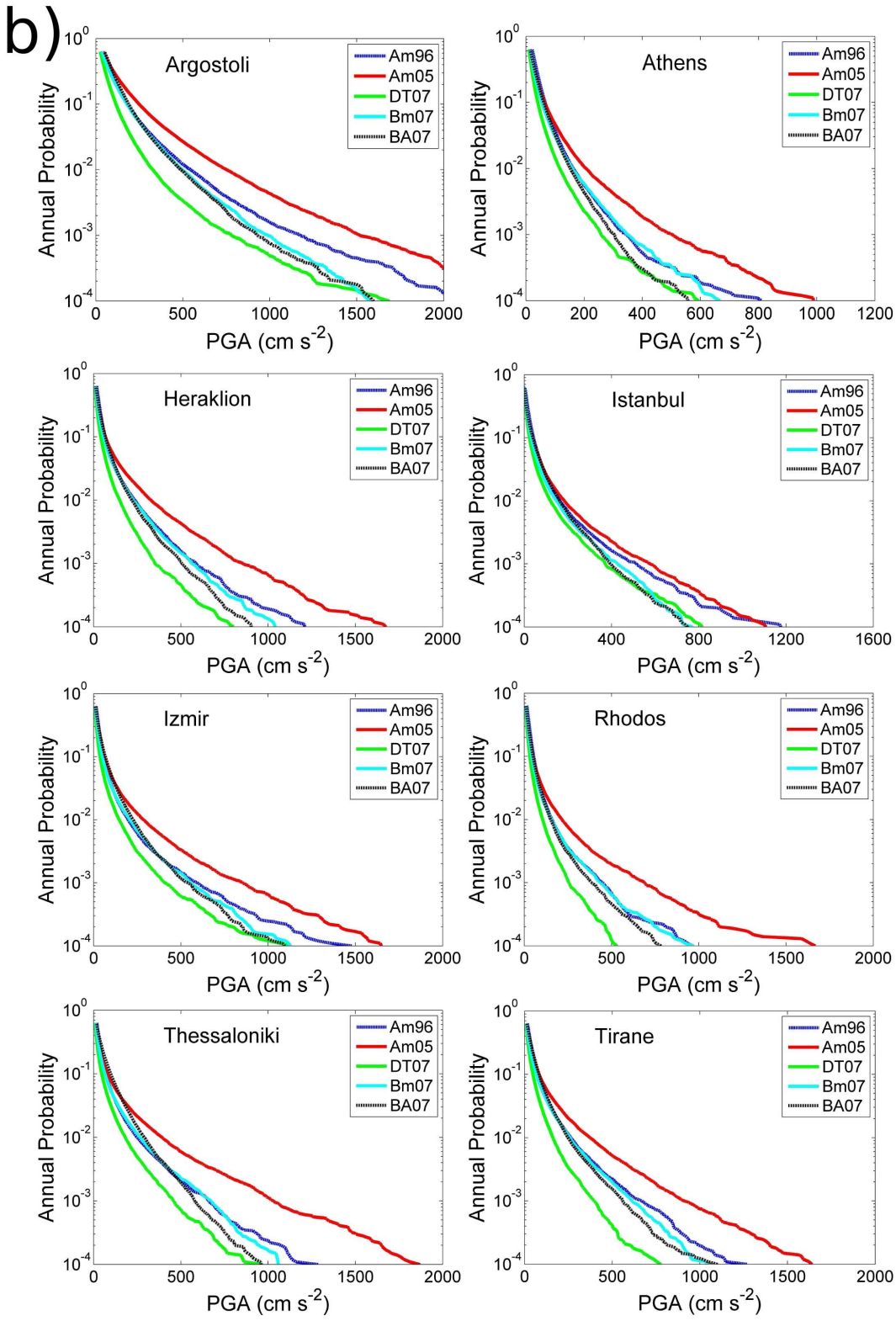

Figure 6.11: b) Seismic Hazard Curves - PZ1990 Source Model

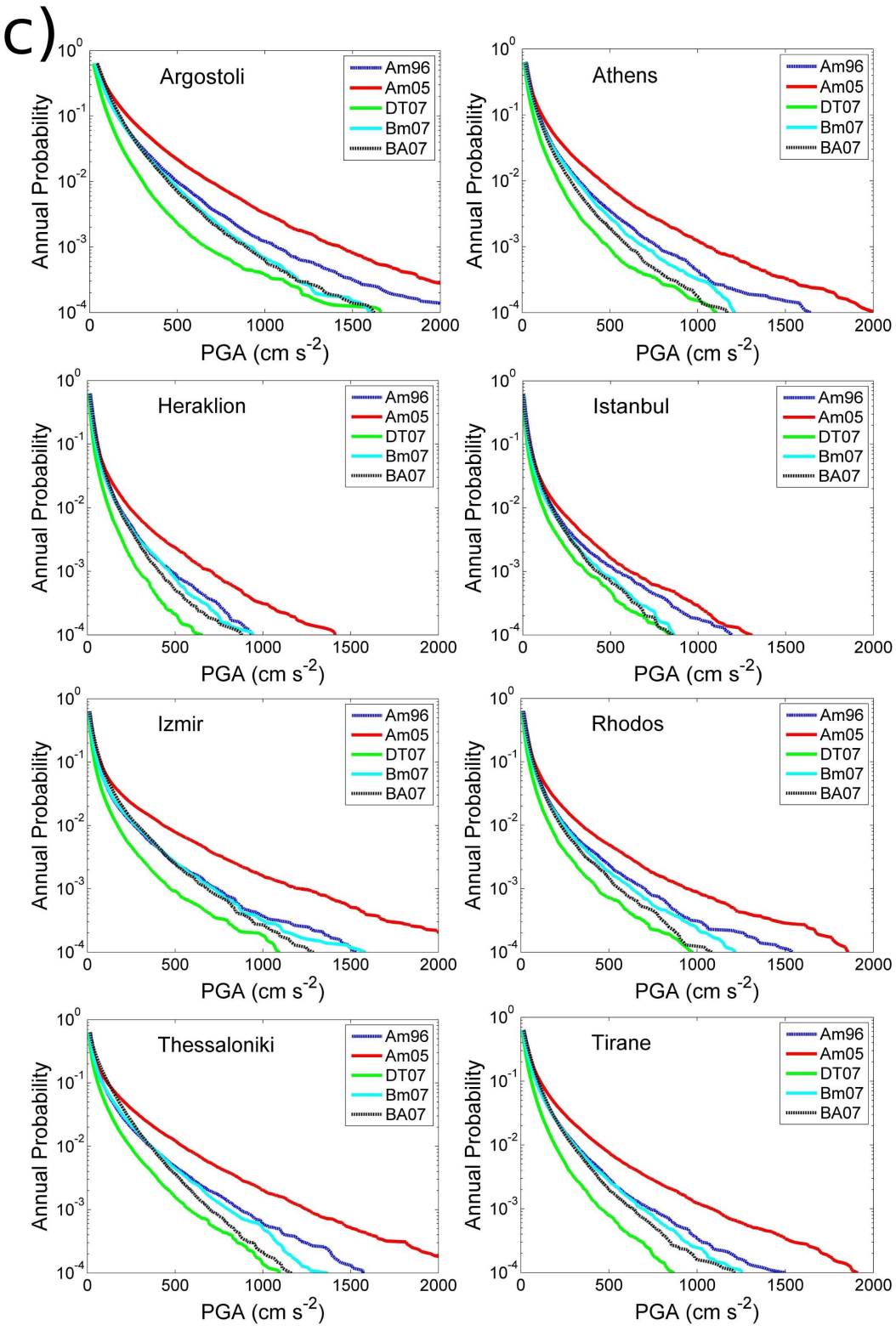

Figure 6.11: c) Seismic Hazard Curves - WT2006 Source Model

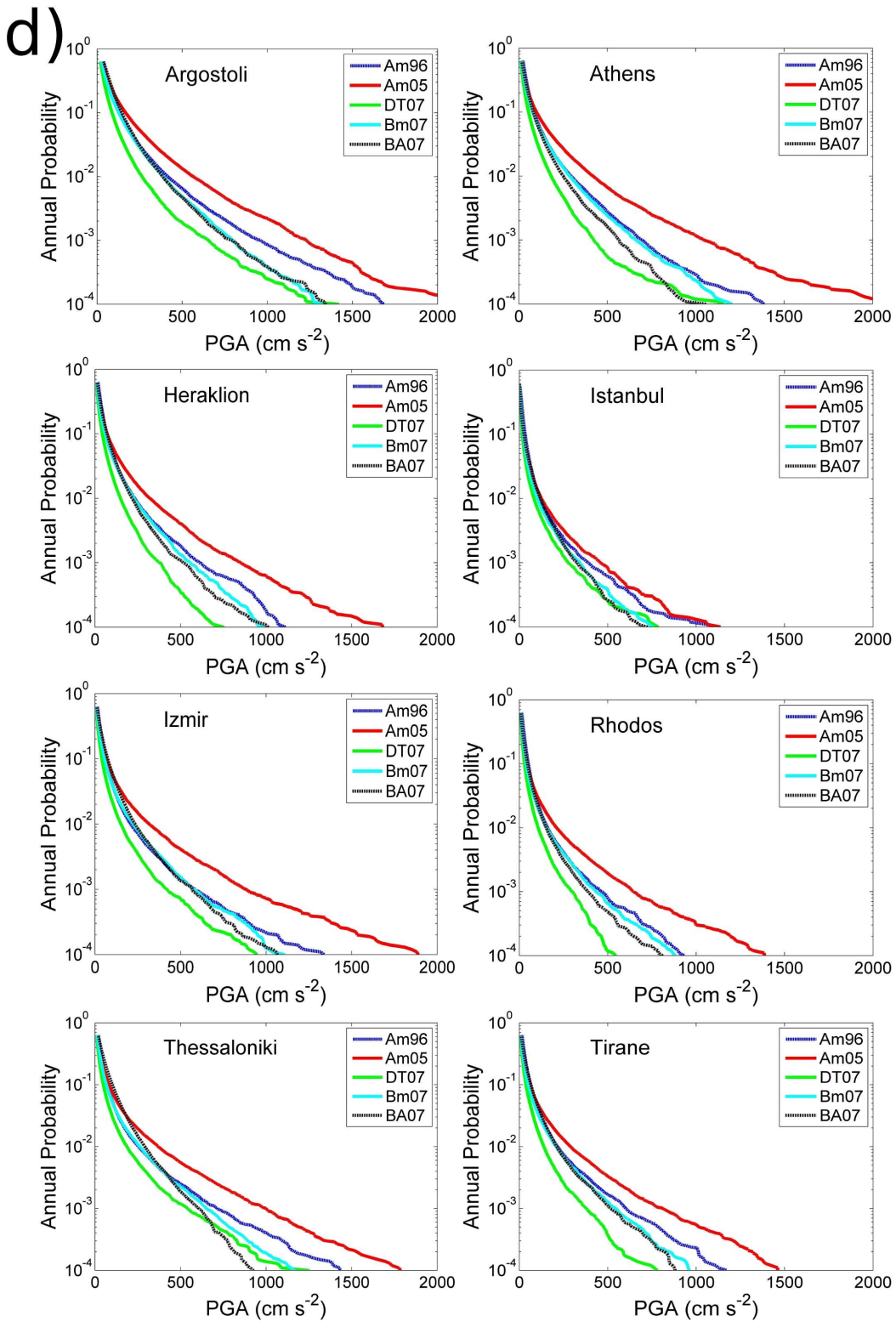

Figure 6.11: d) Seismic Hazard Curves - K27 Source Model

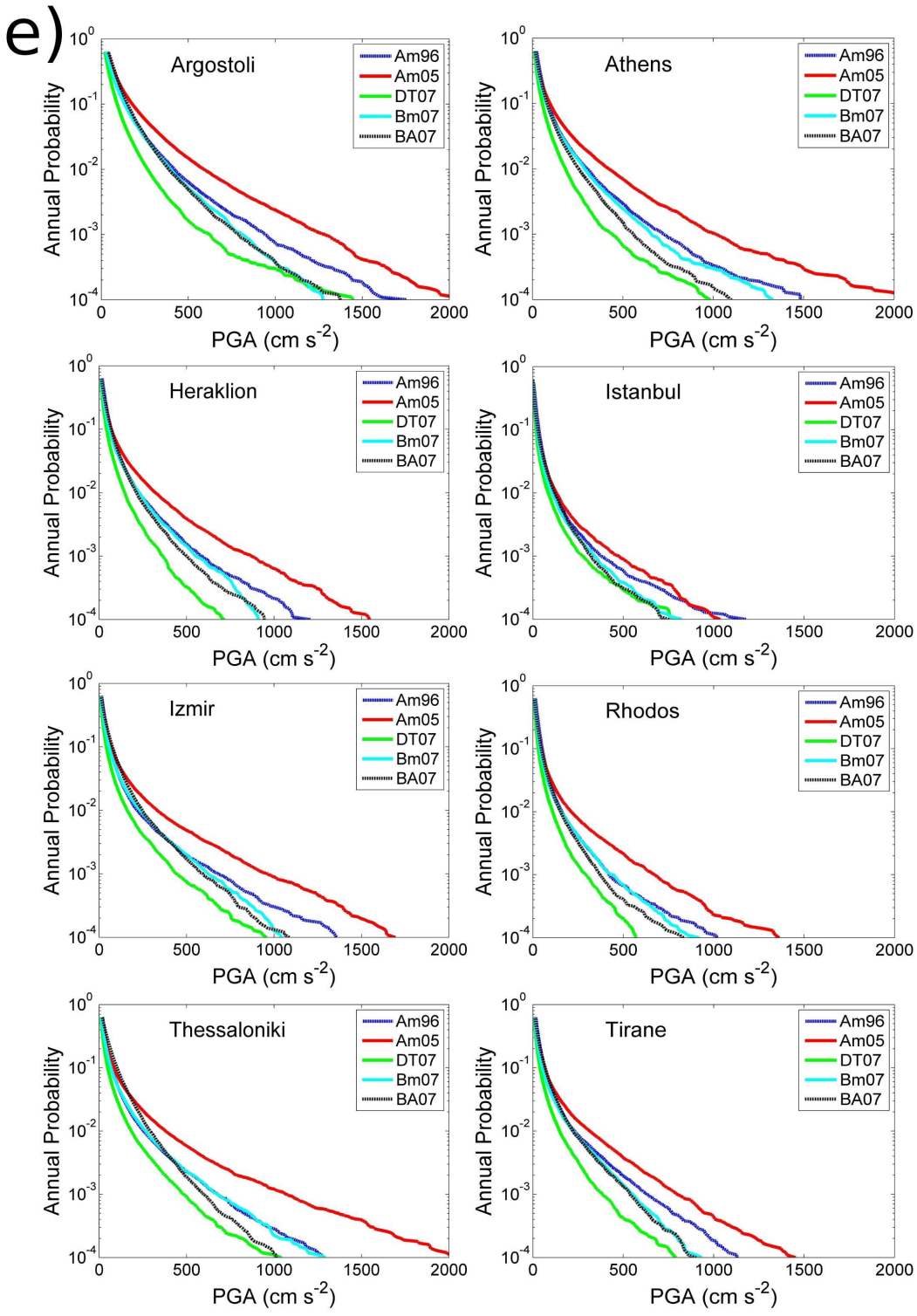

Figure 6.11: e) Seismic Hazard Curves - K29 Source Model

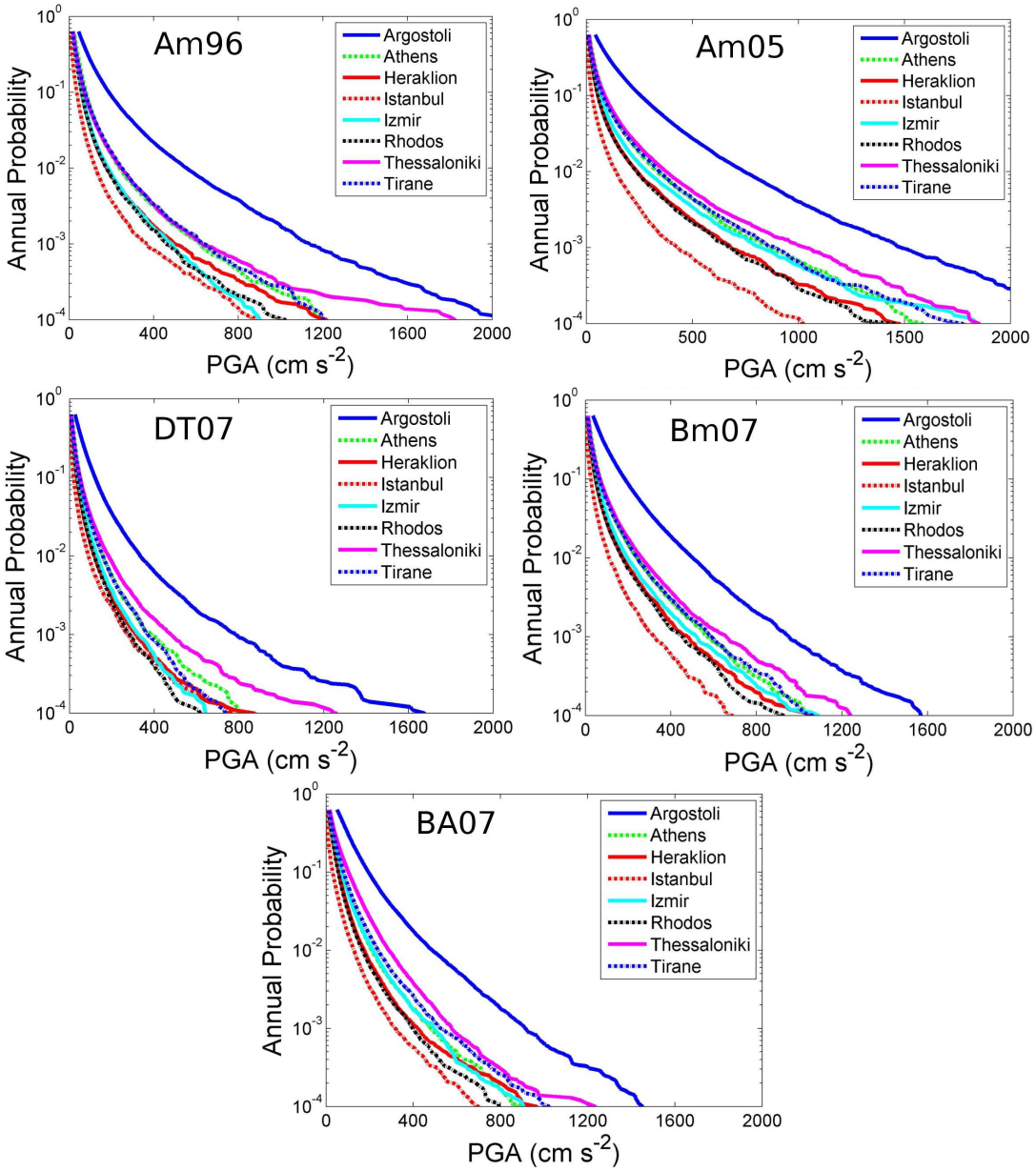

Figure 6.12: Seismic Hazard curves for the 8 sites assuming the PP2000 source model and varying the attenuation model.

Across all sites the DT07 relation predicts lower ground accelerations, whilst Am05 predicts very high accelerations. The curves created using Am96, Bm07 and BA07 show reasonable agreement at higher annual probabilities of being exceeded, but begin to diverge at longer return periods.

The hazard curves for the 8 sites bear out the same trends that were suggested by the hazard map. Argostoli, in the Ionian Islands, has the highest level of hazard, as illustrated in the maps and clearly demonstrated in Figure 6.12. The main impact that the choice of attenuation model has is on the divergence of the hazard curves at low annual probabilities of being exceeded.

Increasing roughness (noise) in the hazard curves at low annual probabilities is a well-established phenomenon associated with the Monte Carlo technique. At these probabilities hazard comes from the largest events which are found less frequently in the synthetic catalogue. The noise suggests that the synthetic catalogues are not sufficiently long enough to sample the largest events in the catalogues (Musson, 2000). This can be rectified by increasing the synthetic catalogue length. Since these curves are derived from an Aegean-wide synthetic catalogue, increasing duration comes at a considerable computational cost. For a single site specific analysis it is more prudent to consider only the area considered relevant for seismic hazard at the site, then simulating longer duration synthetic catalogues. For the purposes of this analysis these curves are sufficiently robust to estimate hazard for annual probabilities as low as 10^{-3} (return period of 1,000 years), and give a good indication of the likely ground motion with a 10^{-4} annual probability of being exceeded (return period of 10,000 years).

Are these hazard curves good indicators of the true seismic hazard at a site? Are the accelerations predicted too high or too low? Without strong motion records extending over the full return period of interest, it is not possible to answer this question directly. There may, however, be some proxies that can infer the range of ground motions that may have been observed at a site. Consider Argostoli for example. The hazard curves for the DT07, Bm07 and BA07 relations suggest that the PGA with a 10^{-4} annual probability of being exceeded is close to 1,200 $cm\ s^{-2}$. This is obviously a very high level of PGA so it is not unreasonable to question whether such a level of acceleration can be observed at that site. Whilst such high accelerations have not been seen in the observed strong motion records, there is evidence from historical macroseismic intensities that such ground accelerations are feasible.

The 1953 Cephalonia earthquake (M_W 7.1), which produced an I_{MM} of X at Argostoli (Papazachos *et al.*, 1997), can be used as an example. Using the Wald *et al.* (1999) empirical relation between I_{MM} and PGA, PGAs around 1200 $cm\ s^{-2}$ would produce likely intensities of IX to X. The Tselentis and Danciu (2008) empirical relation puts this figure at X. Although both relations are being extrapolated beyond their derived range, this observation implies that such a PGA is feasible at Argostoli. At the same site MM intensities of VII and VIII have been observed several times in the last century from smaller earthquakes. These lower intensities correspond to PGAs in the range 200 to 650 $cm\ s^{-2}$ (Wald *et al.*, 1999). The likely return periods for this range of PGA are of the order of 100 to 300 years. This is consistent with all but the Am05 curve.

Whilst these curves may be in reasonable agreement with observed macroseismic intensities in Argostoli, this is not necessarily the case elsewhere. The hazard curves are surprisingly low for Istanbul, with the PGA with a 10,000 year return period of approximately 750 to 800 $cm\ s^{-2}$. In the Istanbul earthquake of 1509 (M_W 7.4) intensities as high as I_{MM} X have been determined (Papazachos *et al.*, 1999). It is possible that for Istanbul the 1509 event was perhaps one with a longer return period than 10,000 years. However,

intensities in the range VIII to XI have been observed at various sites across the Marmara region many times in the last 1,000 years (several in the last 100 years). Given that the hazard maps indicate that much of the Marmara region has a similar hazard to Istanbul then it is likely these maps are underestimating hazard in this region.

Historical data such as macroseismic intensity can only reasonably constrain the hazard at higher annual probabilities. When considering return periods of many thousands of years, there is little information available to constrain the hazard estimates. Recent studies have used precariously balanced rocks (PBRs) to provide upper limits to seismic hazard in various parts of the world (Brune, 1996; Anooshehpoor *et al.*, 2004; Schurch and Becker, 2005; Stirling and Anooshehpoor, 2006). This is certainly a useful method for identifying upper limits on ground motion levels observed during a time period. They do not necessarily give any solid indication as to how often lesser ground motions have been observed. No studies have applied this technique to the Aegean region to date, so no such constraints are available here.

Figure 6.11 shows the impact that the choice of attenuation relation has on the hazard curve for a particular site. For comparison, the impact of the choice of source model on the hazard curve for the same sites is shown in Figure 6.13. The BA07 curve is used for each site and, as before, faulting is considered to be normal or strike-slip/normal where defined. A comparison of the hazard curves for 8 sites (also using BA07 attenuation) is shown in Figure 6.14. Generally there is less variability in the hazard curves when different models are considered, although there are some exceptions. For most sites the WT2006 model produces the hazard curve with the highest accelerations for a given annual probability of being exceeded. The K-means derived models (K27 and K29) give the lowest. The PZ1990 source model produces a much lower seismic hazard curve for Athens than the other models do.

The comparisons of the hazard curves for each site (Figure 6.14) further illustrate the impact of the source model on the hazard analysis. Whereas the choice of attenuation model had little impact on the relative difference in hazard for each site, only elucidating higher hazard sites from the lower ones, different source models appear to have a significant impact on the relative hazard. As with the previous results, Argostoli has a consistently higher hazard curve, even when different source models are used. Likewise Istanbul appears to have a lower curve. The relative hazard for the remaining sites changes significantly depending on source model. Clearly, for these sites there are certain aspects of the way in which the seismic sources are modelled that has a significant impact on seismic hazard.

A clear example of the impact of uncertainty in source model can be seen in Athens. It is only in the PZ1990 and the WT2006 models that Athens sits outside a uniform source zone. Even then, in the WT2006 model the edge of the source zone is very close to the city. Being outside of a source zone means that seismicity is determined by the observed

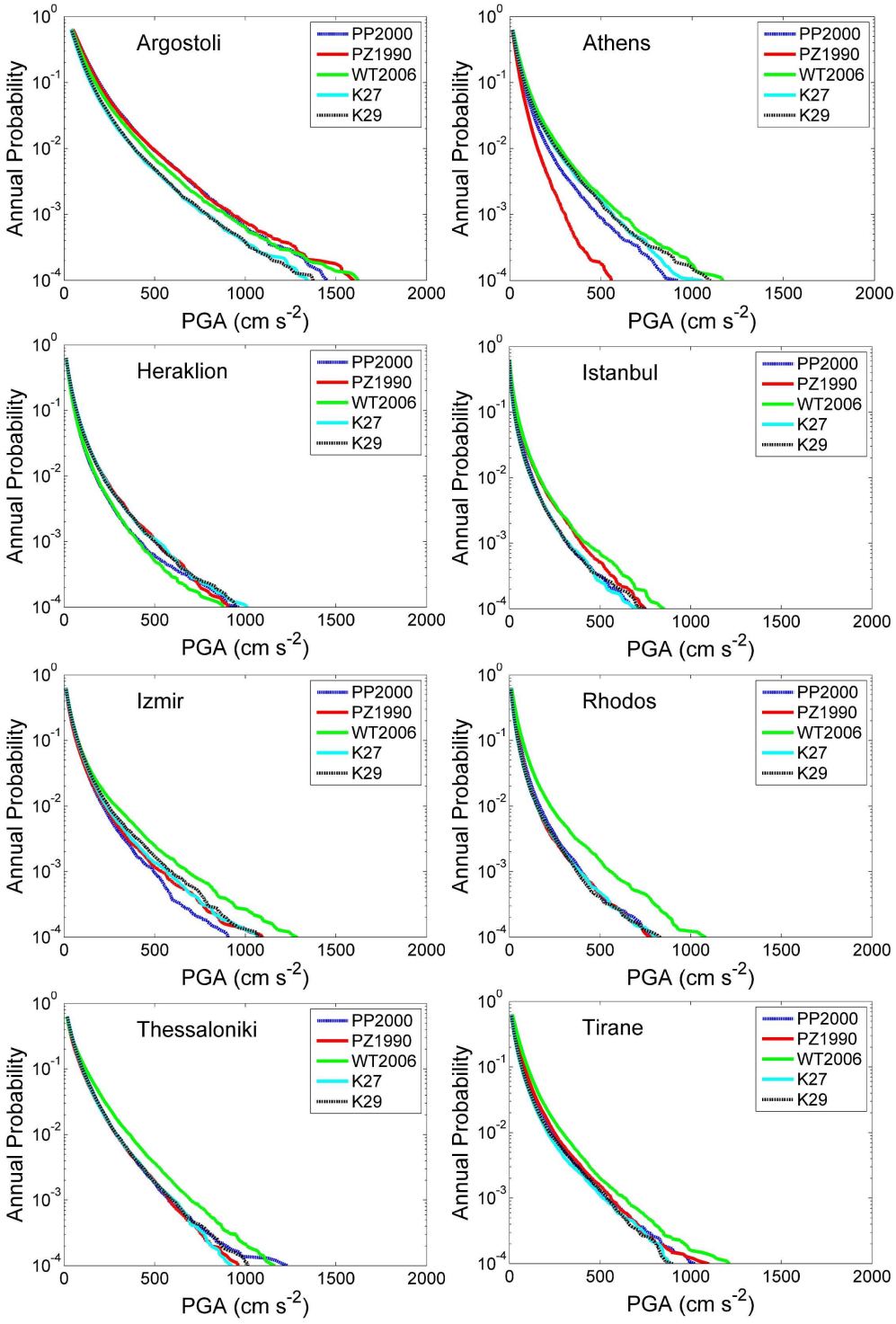

Figure 6.13: Seismic hazard curves using the BA07 attenuation relation and varying the source model.

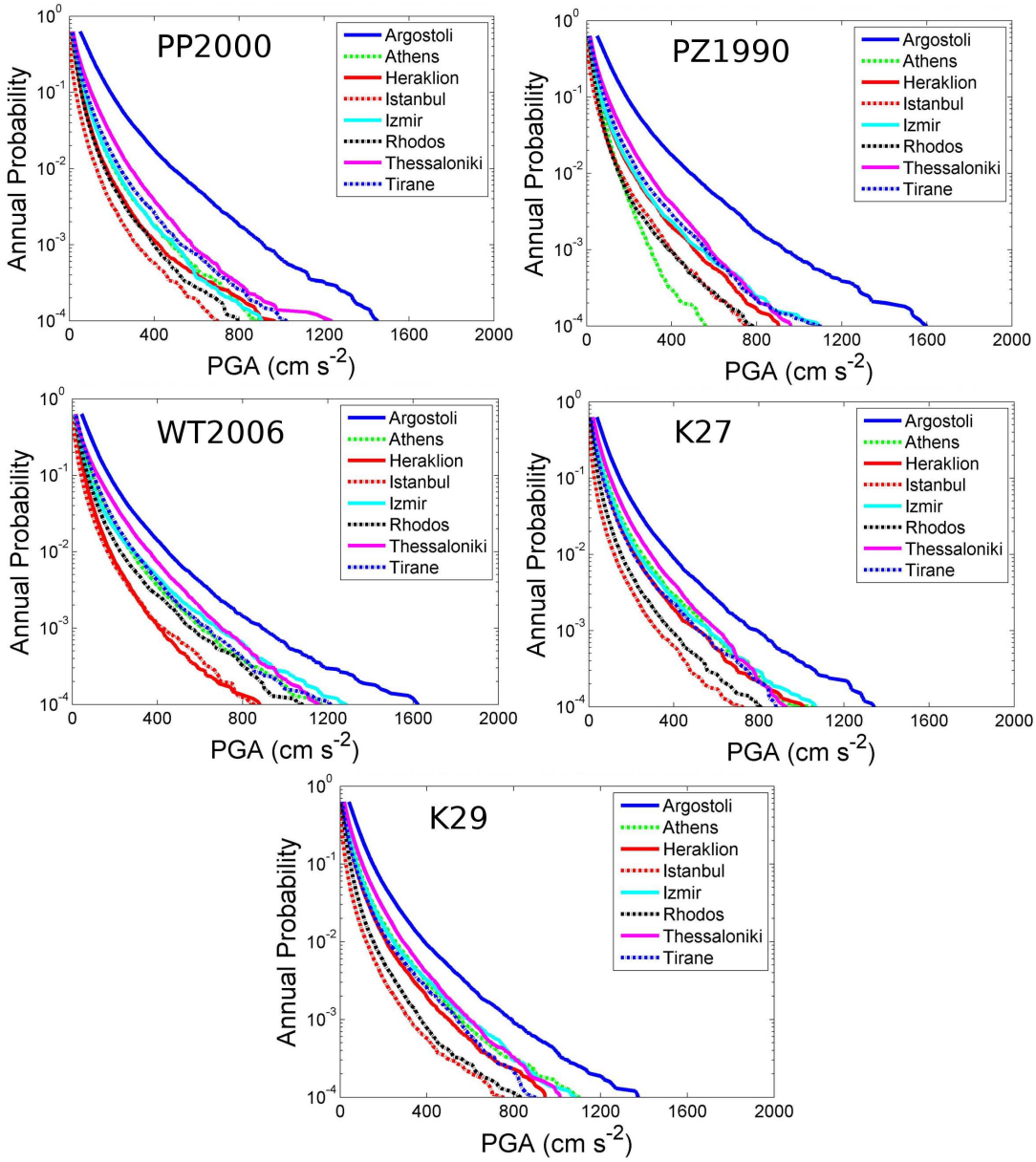

Figure 6.14: Comparison of the Seismic hazard curves for the 8 sites for different source models (BA07 attenuation relation is assumed)

distribution of epicentres. Although there are some notable epicentres around the Athens region, there are very few possible source locations close enough to cause damage; the possible exception being that of the 1999 Athens earthquake.

Despite some disparity in the hazard curves for each source model, the variation is not nearly as obvious as it was for the attenuation relation. This would suggest that the greatest source of epistemic uncertainty in the seismic hazard analysis for this region is the attenuation model, followed closely by the source model. Although the source model may generate maps that appear quite different, for a particular site the greatest differences arise due to the attenuation model. This is a common observation when investigating the uncertainty in seismic hazard analysis.

6.5.3 Uniform Hazard Spectra

It has been shown how the Monte Carlo approach can perform a seismic hazard analysis for a site and across a region. This methodology can also produce another common requirement of a seismic hazard analysis, which is a uniform hazard spectrum (UHS). The spectral acceleration with a P % probability of being exceeded in T years can then be selected independently for each period of acceleration and plotted out in a curve (Abrahamson, 2006). This can be a useful tool for engineering purposes as it can identify the load that a structure may experience within its design life for each spectral period.

The UHS does not correspond to the response spectrum for a specific event. This is a common area of confusion from some of the end users of seismic hazard analysis. As hazard for each period is determined independently, hazard at one end of the UHS is likely to be attributable to different earthquakes from hazard at the other end of the spectrum. Shorter period accelerations may be attributable to smaller near-field sources, whilst longer periods arise form larger sources, possibly further a field.

It has been shown in chapter 5 that the Am96, Am05, DT07, Bm07 and BA07 attenuation relations all contain spectral ordinates, thus making them suitable for this purpose. For the eight sites considered previously, the UHS with a 10 % probability of being exceeded in 50 years is shown in Figure 6.15. As each attenuation relation has a slightly different set of spectral ordinates, the UHS is therefore linearly interpolated to a reference set of ordinates in the range $T = 0.1,\ 0.2,\ 0.3 \ldots 4$. It must also be noted that the Bm07 relation only considered periods up to T = 0.5. Extrapolation beyond this range results in an artificial curve. This is shown for comparison, but it must be recognised that the Bm07 relation is not valid at periods longer than 0.5 s. Similarly Am96 extends only as far as 2 s, whilst Am05 extends up to 2.5 s. These attenuation relations are also invalid if extrapolated beyond these periods. The period above which extrapolation begins is marked on the relevant curves in Figure 6.15.

Several trends that were observed in the PGA data are also apparent when considering

the UHS. The most obvious difference in these spectra is that the Am05 relation is predicting very high accelerations for shorter periods, whilst DT07 predicts lower accelerations. There is a general convergence in spectra in above T = 0.6 s. This would strongly suggest that the disparity between the relations arises due to the difference in modelling accelerations from smaller near-field events. Generally, there is better agreement between the curves when considering larger earthquakes at greater distances.

The levels spectral acceleration seen in the UHS for Argostoli are clearly very large. Most of the UHS seem to predict accelerations around 1500 $cm\ s^{-2}$ at T = 0.1 s, rising to a peak between 1700 and 2000 $cm\ s^{-2}$ in the T = 0.2 s to 0.3 s range before eventually decaying to accelerations less that 200 $cm\ s^{-2}$ at longer periods. Clearly the curvature of the Bm07 spectrum above T = 0.5 s is an artefact of the extrapolation. As with the PGA relations there is better agreement between Am96, Bm07 and BA07 in the 0.1 s to 0.5 s spectral range.

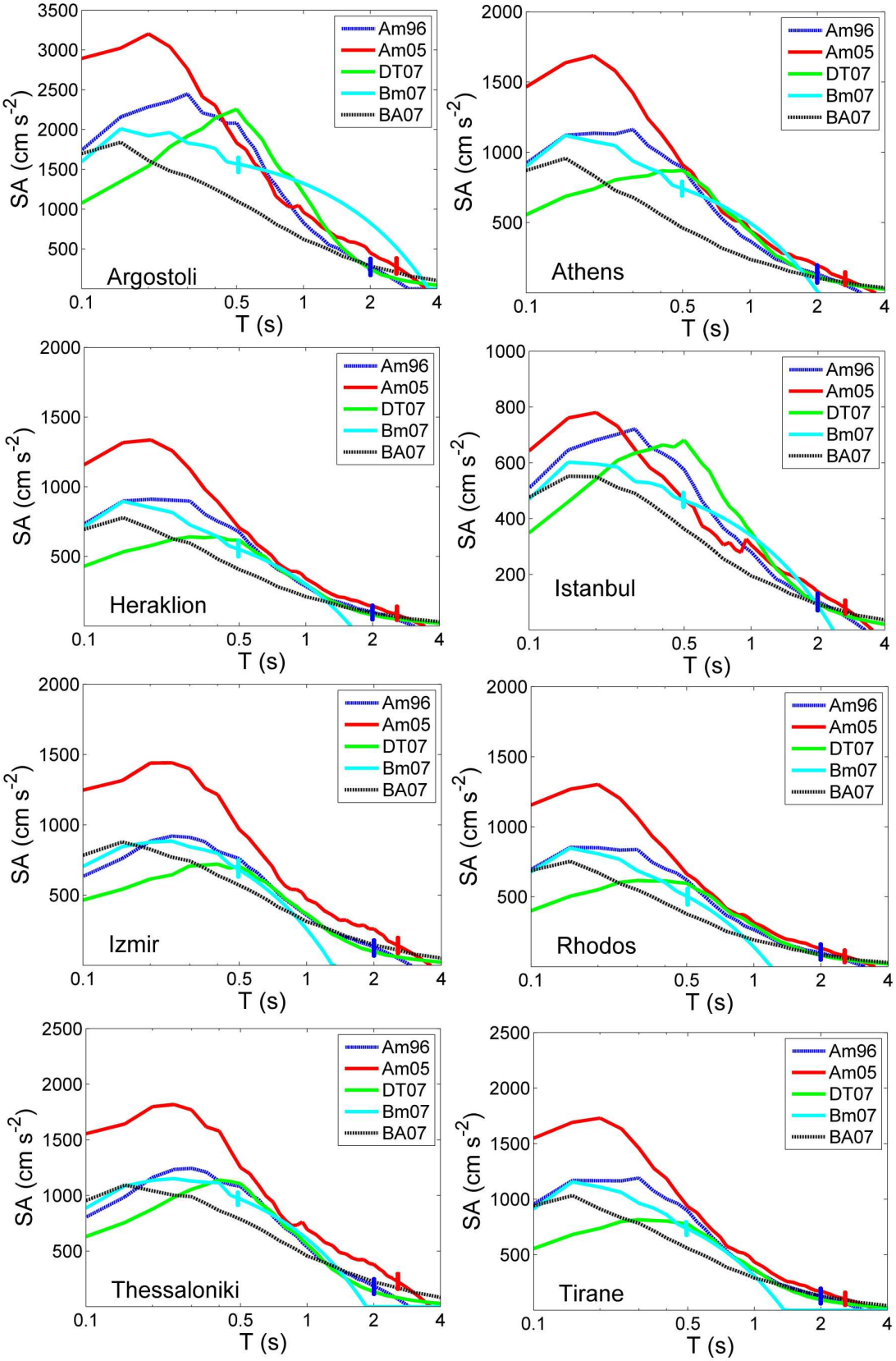

Figure 6.15: a) UHS with a 10 % probability of being exceeded in 50 years - PP2000. Vertical bars on the Bm07, Am96 and Am05 curves indicate the point above which the UHS is extrapolated.

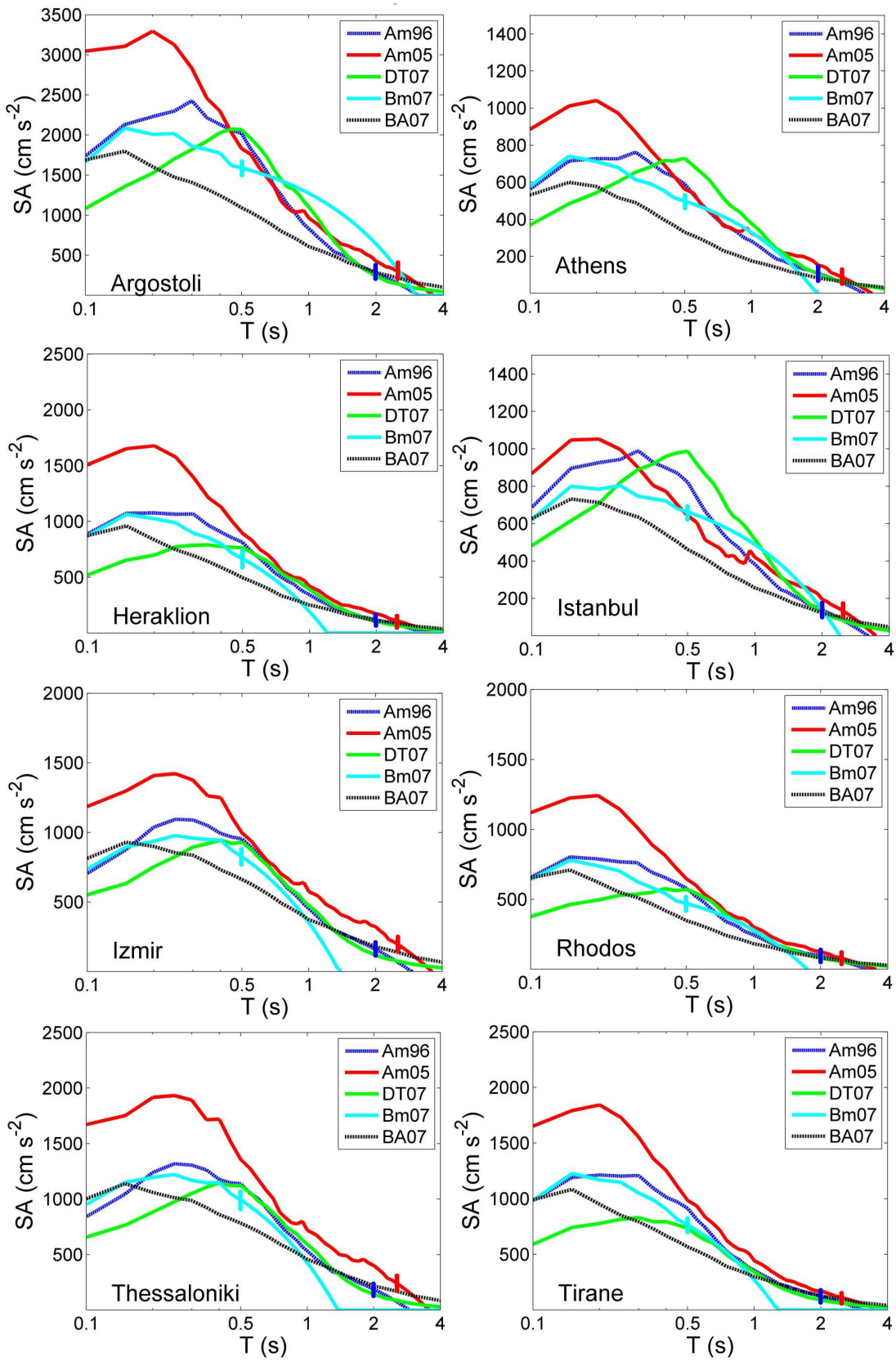

Figure 6.15: b) UHS with a 10 % probability of being exceeded in 50 years - PZ1990. Vertical bars on the Bm07, Am96 and Am05 curves indicate the point above which the UHS is extrapolated.

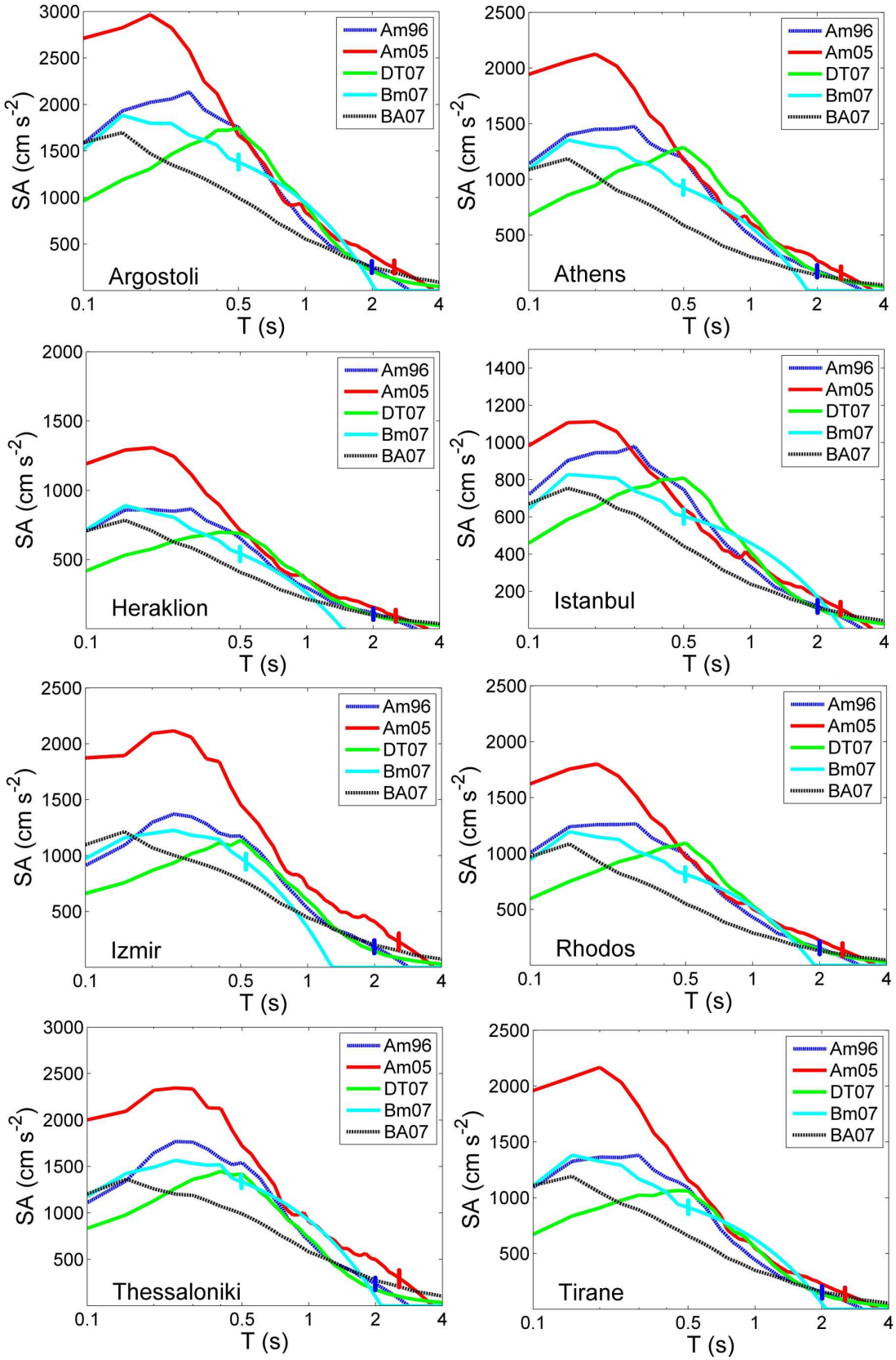

Figure 6.15: c) UHS with a 10 % probability of being exceeded in 50 years - WT2006. Vertical bars on the Bm07, Am96 and Am05 curves indicate the point above which the UHS is extrapolated.

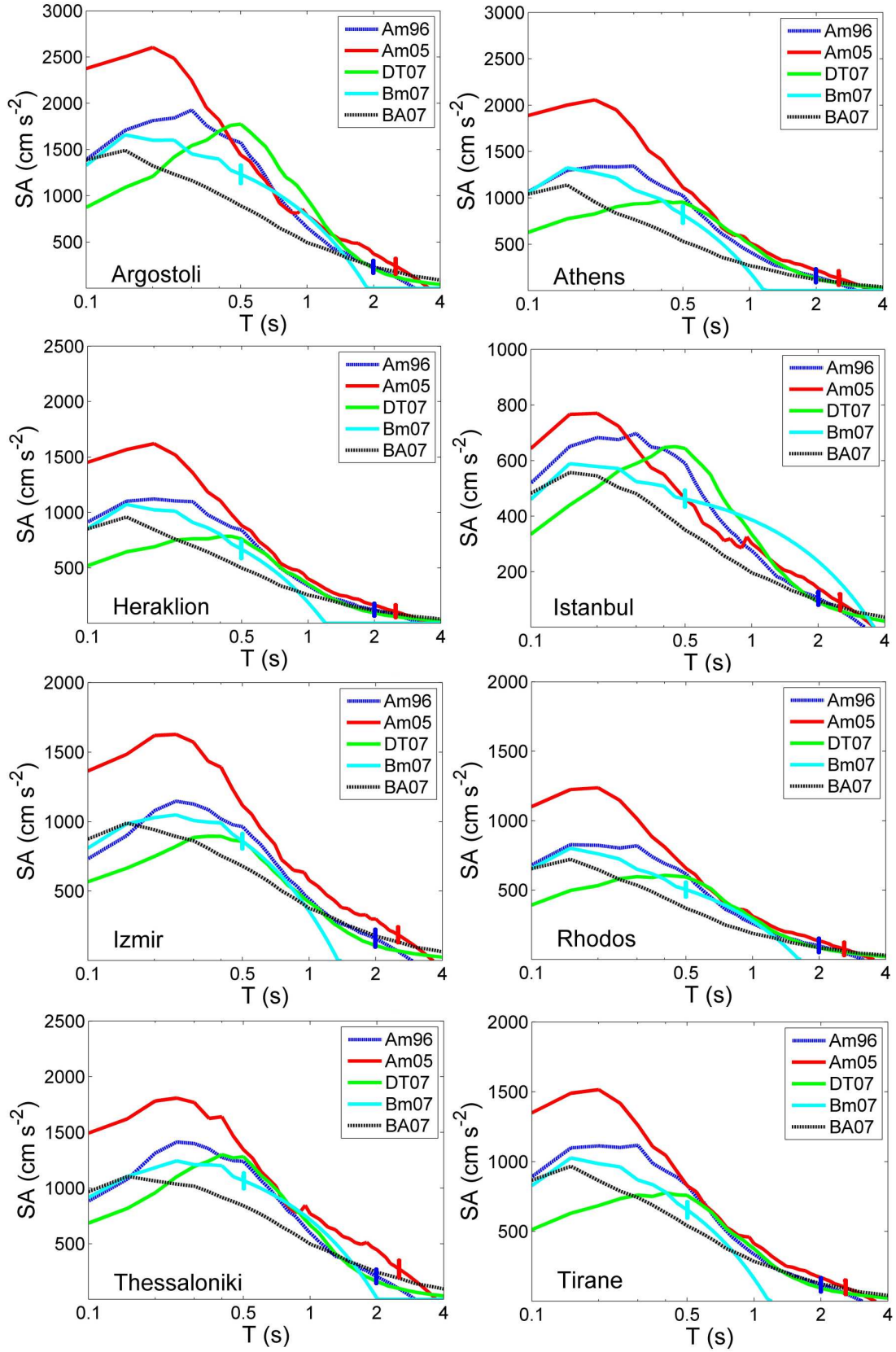

Figure 6.15: d) UHS with a 10 % probability of being exceeded in 50 years - K27. Vertical bars on the Bm07, Am96 and Am05 curves indicate the point above which the UHS is extrapolated.

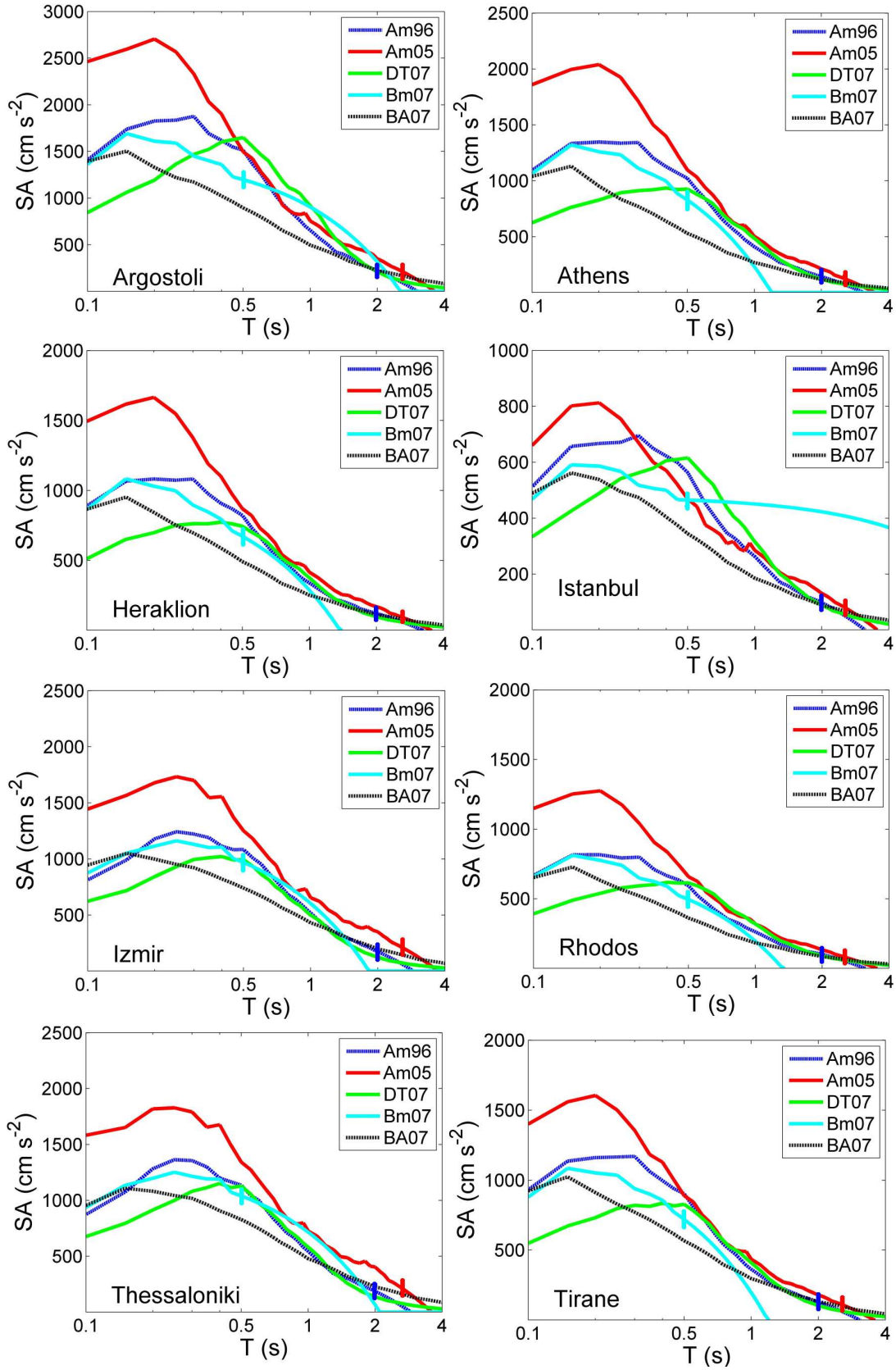

Figure 6.15: e) UHS with a 10 % probability of being exceeded in 50 years - K29. Vertical bars on the Bm07, Am96 and Am05 curves indicate the point above which the UHS is extrapolated.

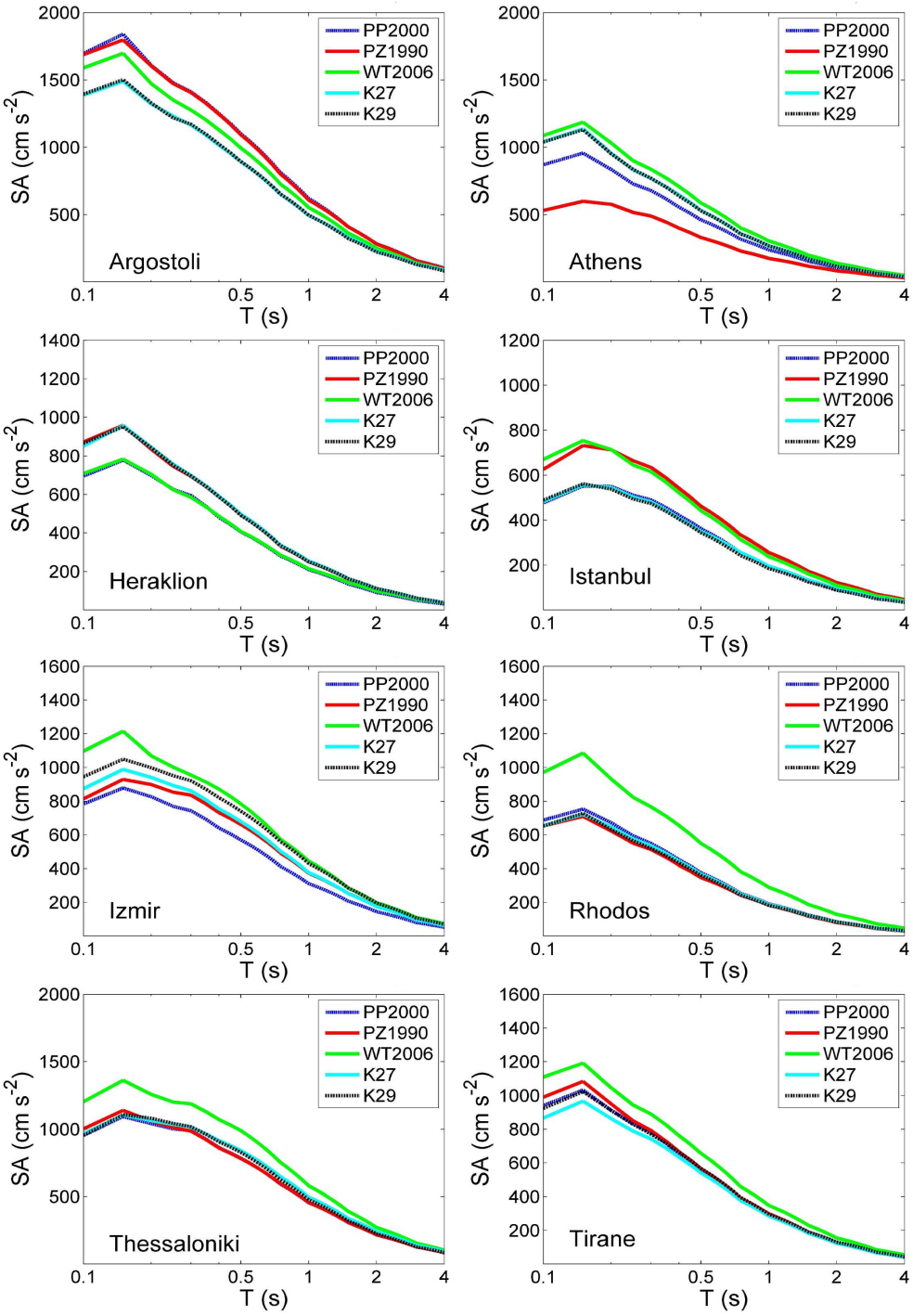

Figure 6.16: Comparison of UHS with a 10 % probability of being exceeded in 50 years with different source model (BA07 attenuation relation assumed)

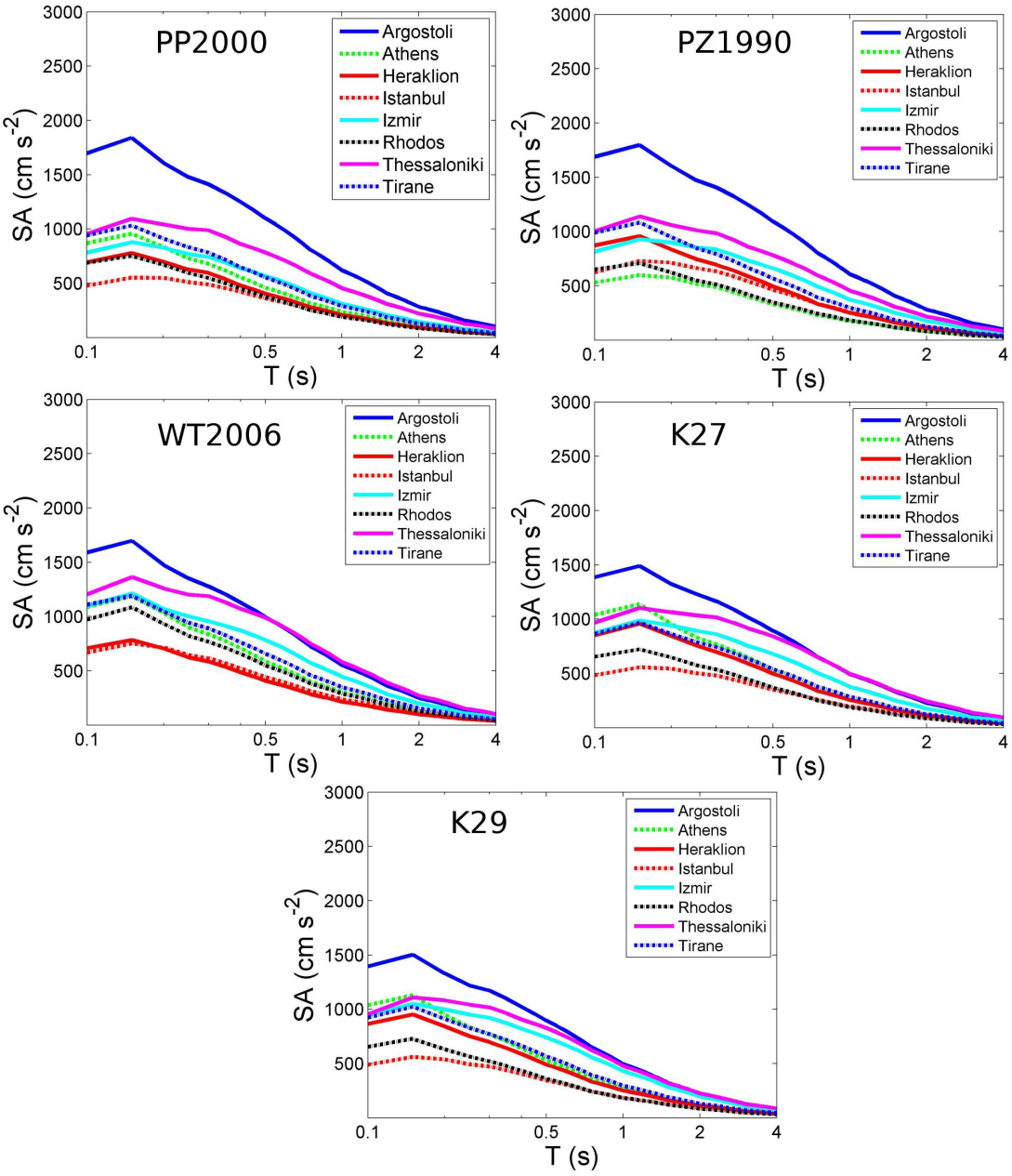

Figure 6.17: Comparison of UHS (for a 10 % probability of being exceeded in 50 years) for 8 cities, assuming the BA07 attenuation and varying the source model

Looking at the variation in UHS, it is not obvious that hazard in one site is controlled by a different magnitude of earthquake than for other sites. This is inconsistent with the seismotectonics of the Aegean region. Larger earthquakes have been observed in the Hellenic arc region than for Northern Greece for example. The same is true for the Marmara region. Differences in rate and recurrence period may explain why PGA hazard appears lower in Heraklion and Istanbul than historical seismicity would imply. If hazard at these sites is more likely to arise from large infrequent events at moderate distances, then it is not unreasonable to expect that the UHS should show larger accelerations at longer periods for these sites than for other locations in the Aegean. This does not appear to be the case. The question arises as to whether this deficiency is due to the source model,

the attenuation model or the Monte Carlo process in general. Figure 6.16 shows the UHS for each site, this time using the BA07 relation but with the source model compared.

The uniform hazard spectra shown in Figure 6.16 do indicate that the source model has an influence on the shape of the curve, albeit less significant that the differences in attenuation relation. It would appear that for most sites the WT2006 model predicts higher accelerations across the spectral range than most of the others. Notable exceptions to this being Argostoli and Heraklion. In most locations the K27 and K29 models produce near identical curves. This suggests a consistency in partition between the two models over much of the Aegean region. Direct comparison of the UHS for each of the 8 sites, for each different source model, is also given in Figure 6.17.

Most of the variation in spectral acceleration is seen in the higher frequency end of the spectrum. The only apparent difference in UHS shape is that a "bulge" of higher accelerations can be seen in the 0.3 s to 0.6 s range. This is clearly visible in the UHS for Thessaloniki and Izmir, and can be seen to a lesser extent in the Tirane and Rhodos UHS (Figure 6.17). It may be possible to make inferences as to whether these bulges suggest that larger, longer distance earthquakes are controlling hazard for these spectral periods. The results cannot confirm this.

6.6 Disaggregation

The UHS does not correspond to a single earthquake scenario, as each period of acceleration is treated independently. For engineering design this can be problematic. Hazard at shorter periods may be due to smaller near field earthquakes, whilst at long periods larger distant earthquakes may come to dominate. For the purposes of engineering design (and even seismic risk analysis) it is extremely useful to be able to identify a scenario or small number of scenarios that represent the most likely event(s) to cause damage at a site. Disaggregation is used to identify the contribution of specific earthquake magnitude and distance combinations to hazard at a site. It is now a standard feature in seismic hazard analysis when using the Cornell (1968) - McGuire (1976) method of PSHA (Bazzurro and Cornell, 1999; Abrahamson, 2006).

A disaggregation simply plots the contribution to the probabilistic hazard analysis of each magnitude and distance combination. Whilst it is not possible to construct such a plot using the Monte Carlo technique, the synthetic earthquake catalogues provide the equivalent information. This is well illustrated in Musson (1999a). For a given site the PGA (or SA) with a probability P of being exceeded in T years is identified. The synthetic catalogue is then searched to find all the events that produce the specified level of PGA at the site, within a given tolerance (here $\pm 10\ cm\ s^{-2}$). By counting the number of events occurring within specified magnitude and distance bins, the relative contribution to hazard from each bin is seen. The "modal" magnitude and distance value can be used

to form the basis of a scenario event for hazard at a site.

The Musson (1999a) disaggregation method is used here to elucidate the sources contributing to hazard at several of the sites that have been shown previously: Argostoli, Athens, Istanbul and Thessaloniki. The results, using the PP2000 source model and BA07 attenuation relation, are shown in Figure 6.18.

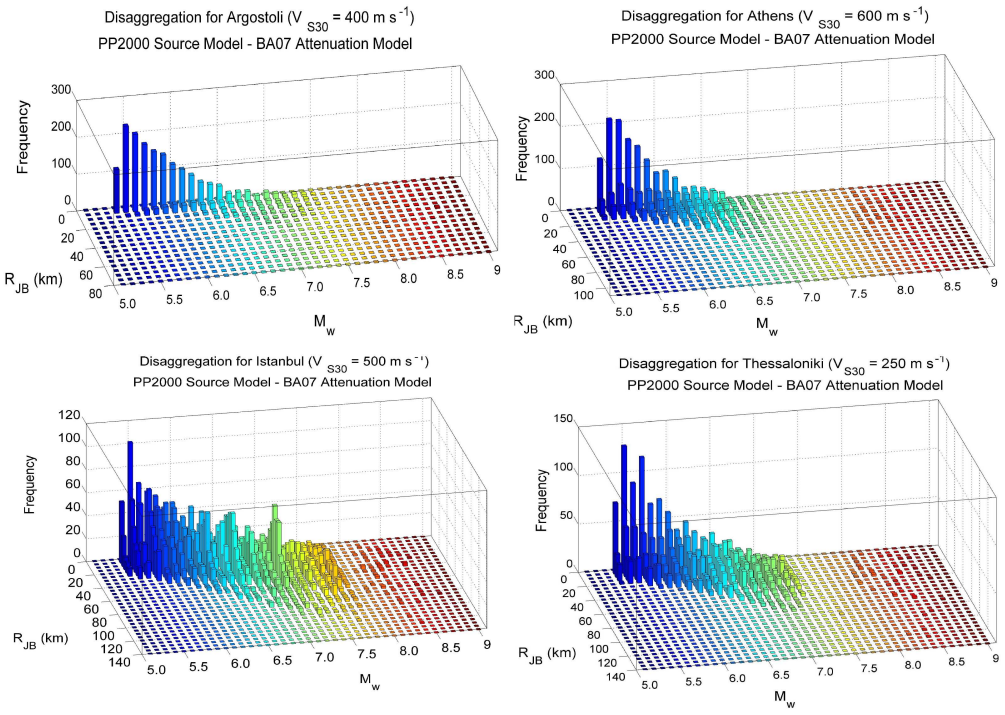

Figure 6.18: Disaggregations for the PGA with a 10 % probability of being exceeded in 50 years for Argostoli, Athens, Istanbul and Thessaloniki. PP2000 source model and BA07 attenuation model assumed.

These disaggregation plots demonstrate some interesting trends with respect to the magnitude and distance of the most likely design earthquake. The result for Argostoli clearly demonstrates that it is small earthquakes over very short Joyner-Boore distances that appear to contribute most significantly to PGA hazard at this return period (1 in 475 years). This illustrates how activity rate can influence the disaggregation and skew the design earthquake due to aleatory variability in the attenuation relation. The simulated aleatory uncertainty value for each ground motion calculation was recorded. Earthquakes in the M_W 5 to M_W 5.5 range were producing the input hazard level of PGA (10 % probability of being exceeded in 50 years) when the percentile value was typically 1σ to 3σ above the median. Since the BA07 attenuation relation is being used, which has a fixed sigma; this cannot be attributed to heteroscedastic aleatory variability.

The PP2000 model has an influence as Argostoli lays within the Cephalonia zone, the most active zone in the model. Under the assumption of a spatially uniform zone and a truncated Gutenberg-Richter recurrence relation it is obvious why small earthquakes in

the extreme near-field are so prevalent. The problem is whether it is realistic to consider a small earthquake with ground motion several standard deviations above the median value to be a sensible design scenario. If so, how is this reconciled with the PGA hazard curves and with the knowledge of seismotectonics in the Ionian region?

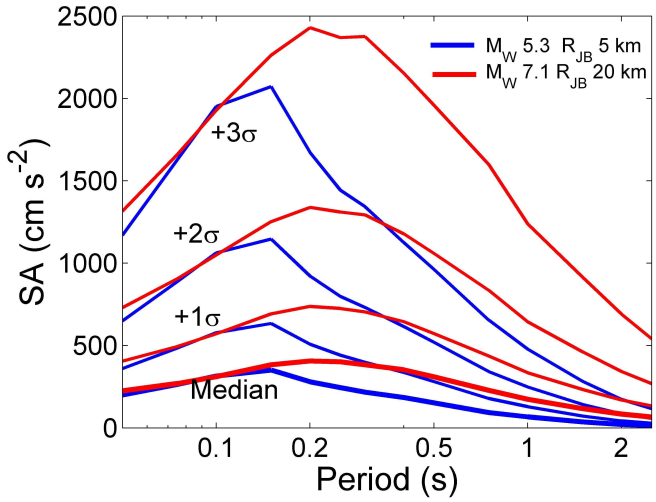

Figure 6.19: Comparison of scenario spectra for a small near-field earthquake (blue) and a larger earthquake at 20 km (red)

A comparison of the spectra for two different earthquake scenarios at Argostoli is given in Figure 6.19. Here the "modal" event from the disaggregation (M_W 5.3, $R_{JB} = 5\ km$) is compared with a scenario based on the 1953 Cephalonia earthquake (M_W 7.1, $R_{JB} = 20\ km$) (Louvari *et al.*, 1999; Papadimitriou, 2002). Even for the median value the PGA is similar, with the Cephalonia event forecasting slightly larger values for the same percentile. For shorter periods of acceleration the two earthquakes follow the same pattern, with the Cephalonia event producing slightly larger accelerations. At periods that are relevant to engineering design (0.2 s to 2 s) the accelerations from the Cephalonia event are much larger. This clearly indicates that for a deterministic scenario the Cephalonia event is more relevant. When taking into consideration recurrence, however, the situation changes. Assuming a power-law recurrence, the modal event is many times more likely than the Cephalonia event. This is borne out by observations of seismicity. If the uncertainty in the distribution of ground motion residuals is independent for each event, then it is likely that for every Cephalonia event observed, there may be several "modal" events with a higher fractile of ground motion, according to this model.

What is the cause of this problem? Is it the source model, the recurrence model, the attenuation model or the Monte Carlo method itself? In the case of the source model, one could throw out the uniform zones assumption and use either the observed distribution of seismicity or the location of the Cephalonia transform fault as a point or plane source. In doing so one may avoid having to consider low magnitude events in the extreme near-field, this is assuming all seismicity is pinned to the fault plane. The problem with this is that observed seismicity is distributed around the fault, and has occurred within a few

kilometres of Argostoli. The source model may be the problem, but similarly a more deterministic source is not realistic given the observed seismicity. The same can be said for the recurrence and attenuation models. In both cases the models are consistent with the observed properties of earthquake behaviour and strong ground motion. Furthermore, Monte Carlo method also allows for the integration of uncertainties in both these models into the hazard calculation. Despite all of this the controlling hazard scenario is unrealistic. This may be considered one of the examples of the inputs and the procedure being correct but still resulting in an incorrect, or at least counter-intuitive, output. It is also clearly showing the impact that aleatory variability in ground motion has on the hazard analysis.

The Argostoli example may not necessarily be the fairest case for this analysis. Located within only a few kilometres of an active fault, in the most active part of the Aegean, the influence of extreme (and possibly ill-constrained) near field events was always going to be a problem. Consider the other examples shown in Figure 6.15: Athens, Thessaloniki and Istanbul. In each of these cases there are a greater proportion of larger distant earthquakes contributing to the hazard at these locations than for Argostoli. Nevertheless, in Athens and Thessaloniki there is still a considerable bias toward smaller events in the extreme near field. For Istanbul, despite the lower level of PGA hazard, the disaggregation is more consistent with the seismotectonics within the region. Here there is a greater proportion of small earthquakes contributing to hazard, but the distribution is becoming more obviously bi-modal with a second peak occurring around M_W 7.1 over distances of 20 to 40 km. This would correspond closely to a scenario event of similar magnitude to the Duzce earthquake (November 1999) occurring along the North Anatolian fault at the eastern end of the Sea of Marmara. Such events have been seen in the historical catalogue. Furthermore, this design event is consistent with a strong ground motion sigma value closer to the median.

Given that the aleatory variability in the strong ground motion attenuation relation appears to be crucial to the choice of design earthquake, it is necessary to consider other attenuation relations. Disaggregations for the same four sites using the PP2000 source model are shown in Figures 6.20 and 6.21. Figure 6.20 is developed using the Bm07 relation, which has a magnitude-dependent variability (increasing variability with lower magnitude). Figure 6.21 is developed using the DT07 relation, which has a fixed aleatory variability term, but lower than BA07.

There is considerable similarity between the disaggregations using the Bm07 and the BA07 relation. For all sites the event of mode is still around M_W 5.5 at a Joyner-Boore distance less than 10 km. As with the previous figure these "modal" events require a high percentile of ground motion (typically above 1σ). There appears to be a slight shift toward larger events over intermediate distances (20 to 40 km). This is accompanied by a drop in the number of large distance events contributing to hazard at a site. This behaviour is attributable to the difference in σ between the two relations. Even at low

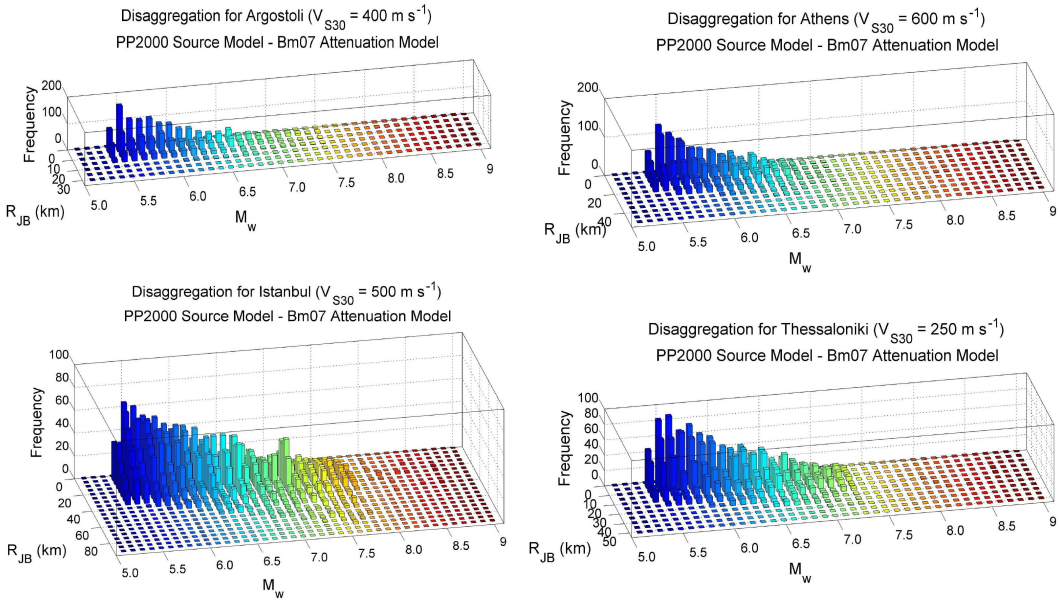

Figure 6.20: Disaggregations for the PGA with a 10 % probability of being exceeded in 50 years for Argostoli, Athens, Istanbul and Thessaloniki. PP2000 source model and Bm07 attenuation model assumed.

magnitudes, σ is lower in the Bm07 relation than the BA07 relation; hence slightly larger events are required to generate the same level of ground motion for the same fractile of the attenuation relation. The σ value for large magnitude events is much lower in the Bm07 relation than the BA07 relation; hence fewer large events contribute to hazard at each site.

Use of the DT07 relation produces a more remarkable change in the shape of the disaggregations, shifting the emphasis toward larger earthquakes. For Argostoli, the design event is now pushed more towards M_W 6.3 to 6.5 at distances of 5 to 10 km. Earthquakes of this magnitude occur regularly along the Cephalonia transform fault making the scenario more feasible. Similarly, for Athens and Thessaloniki, the design earthquakes are tending more toward M_W 6.5 at distances of 20 - 30 km. One interesting development is that for Istanbul the scenario event of a M_W 7.1 earthquake at a distance of 40 km is now the clear "modal" event.

Comparison between disaggregations using DT07 and those using Bm07 and BA07 is a little harder given the properties of the DT07 relation. Firstly the DT07 relation predicts lower PGA for smaller events than either Bm07 or BA07, although the differences are not vast. Secondly, DT07 has a fixed aleatory variability, which is substantially lower than that of BA07 and that of Bm07 for small events. By reducing the overall PGA contribution from small events the DT07 is beginning to shift the design event(s) toward the rarer large earthquakes.

The impact that the choice of attenuation relation has on the type of earthquakes con-

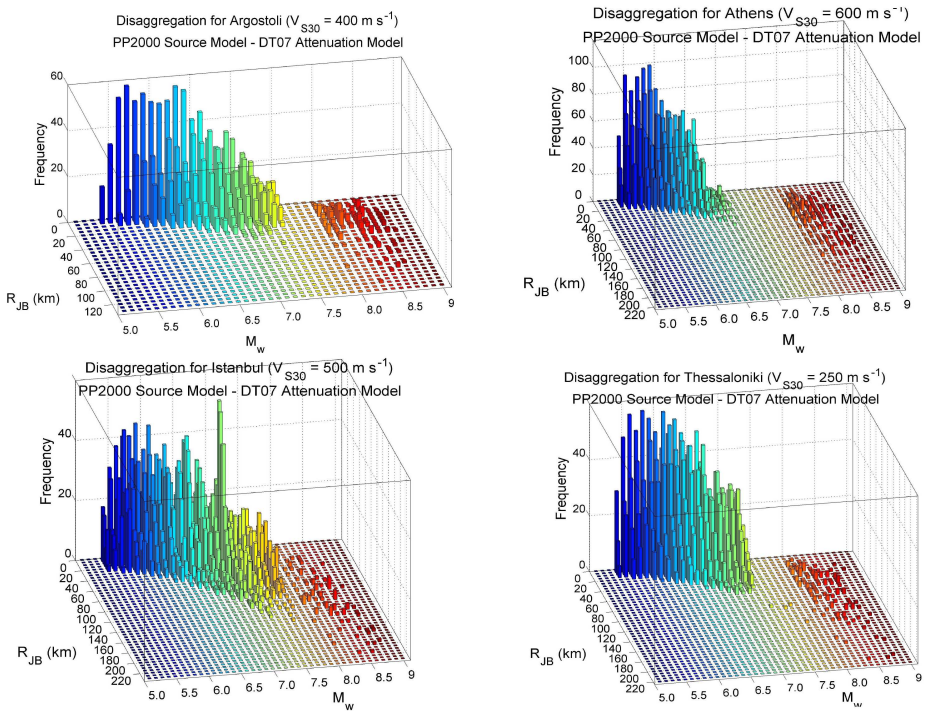

Figure 6.21: Disaggregations for the PGA with a 10 % probability of being exceeded in 50 years for Argostoli, Athens, Istanbul and Thessaloniki. PP2000 source model and DT07 attenuation model assumed.

tributing to hazard at a site is obvious. Does the choice of source model have the same impact? Figure 6.22 shows a disaggregation for the same four sites, this time using the PZ1990 (the attenuation relation is BA07).

There are very few obvious differences in the disaggregations shown when either the PP2000 or PZ1990 models are used. For all sites the "modal" magnitude appears to have risen slightly in the PZ1990 model from M_W 5.4 to M_W 5.7, at Joyner-Boore distances of less than 10 km (with sigma well above 50 %). Otherwise these disaggregations are much the same as those for PP2000. This result is not entirely surprising as there are many similarities between two models in zones close to each of the cities being considered. The slight shift in "modal" magnitude may arise from the fact that in the PZ1990 source model Athens, Istanbul and Thessaloniki sit on the edge of uniform zones thus reducing the number of small earthquakes simulated very near to each site. Finally, disaggregations using the WT2006 and K27 source models are shown in Figures 6.23 and 6.24 respectively.

The disaggregations when the WT2006 model is used do show more significant differences. It would appear that here the use of the WT2006 model has meant that near-source small magnitude earthquakes more commonly contribute to hazard at the site (in the synthetic catalogues) than they did when the PZ1990 model was used. As with the PP2000 model, all four cities are situated within uniform source zones in the WT2006 model. This distributes the location of epicentres more evenly around each site.

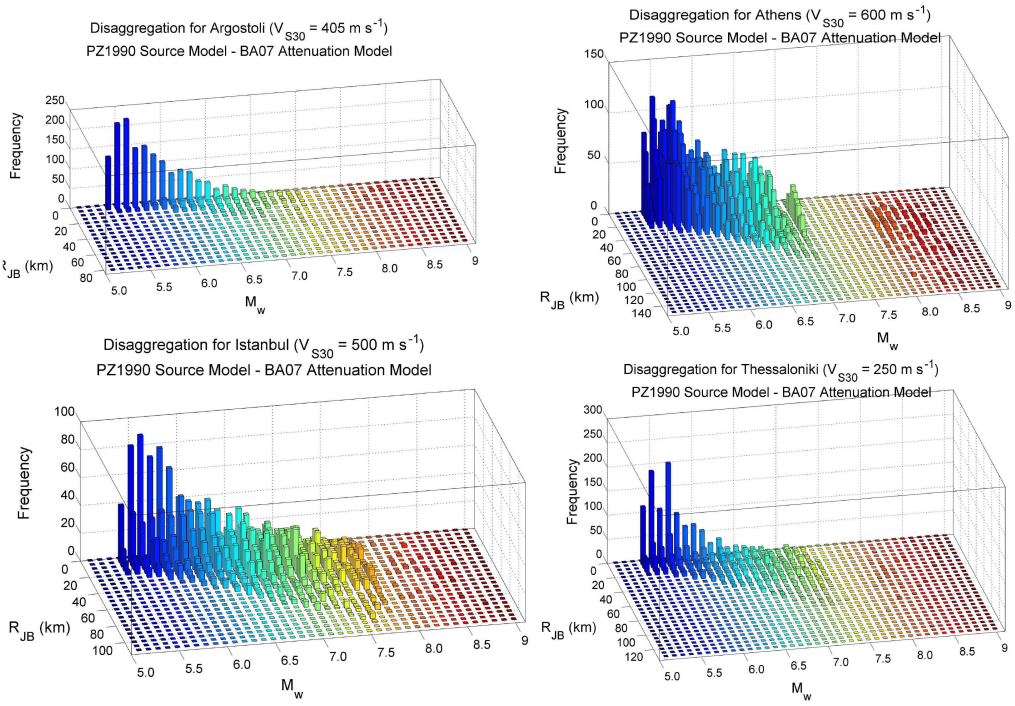

Figure 6.22: Disaggregations for the PGA with a 10 % probability of being exceeded in 50 years for Argostoli, Athens, Istanbul and Thessaloniki. PZ1990 source model and BA07 attenuation model assumed.

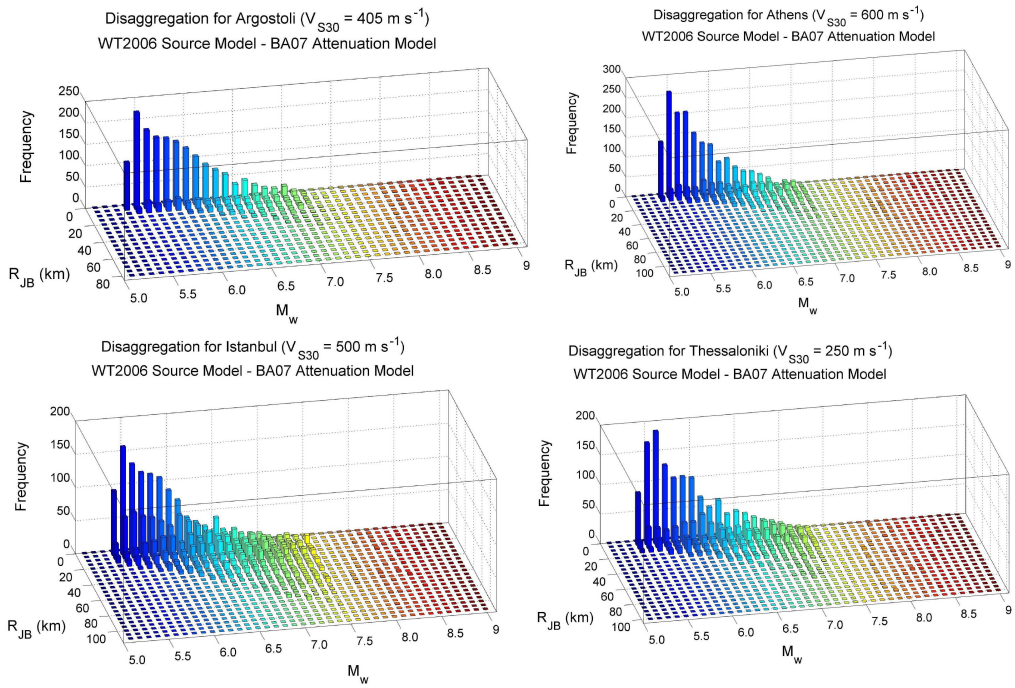

Figure 6.23: Disaggregations for the PGA with a 10 % probability of being exceeded in 50 years for Argostoli, Athens, Istanbul and Thessaloniki. WT2006 source model and BA07 attenuation model assumed.

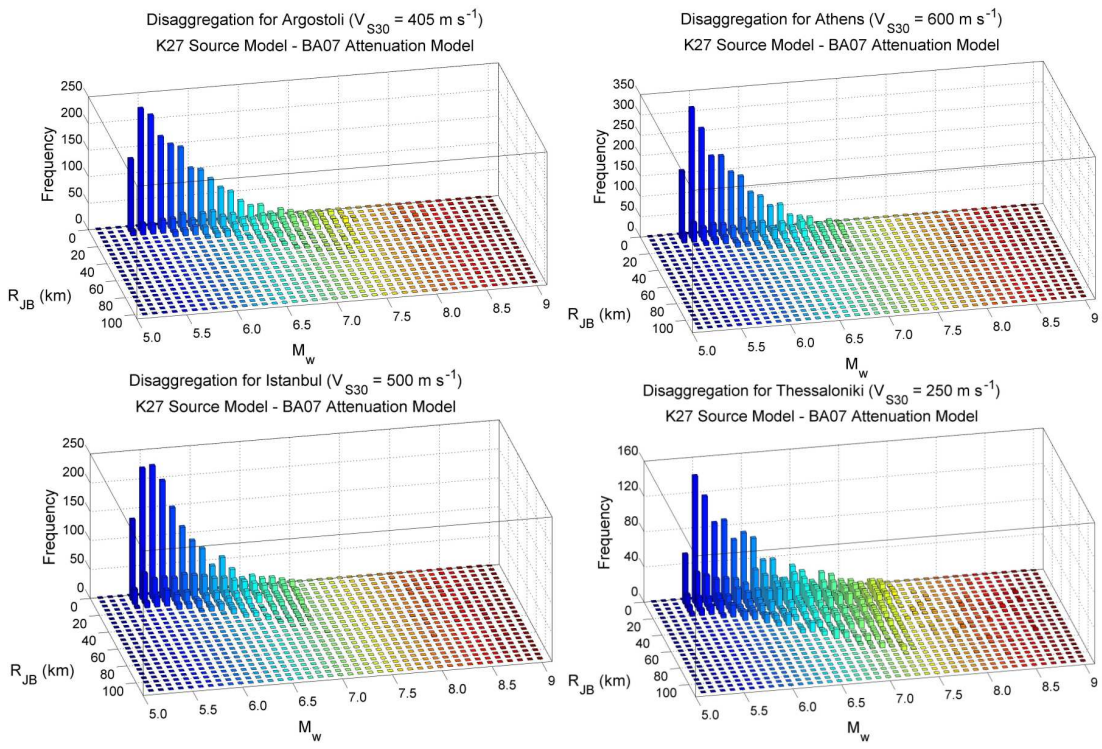

Figure 6.24: Disaggregations for the PGA with a 10 % probability of being exceeded in 50 years for Argostoli, Athens, Istanbul and Thessaloniki. K27 source model and BA07 attenuation model assumed

In contrast, the K27 model appears to have the opposite effect, this time increasing the contribution of near-field, low-magnitude events. This effect is obvious even in the Istanbul disaggregation. Only in Thessaloniki does there appear to be an increase in the number of high magnitude earthquakes contributing to hazard. As with the PP2000 model, the K27 model employs uniform source zones over a wide area, thus increasing the number of possible near-field events contributing to hazard at each city.

These disaggregations reinforce the suggestion that it is attenuation model that contributes most strongly to the uncertainty in seismic hazard at a site. Although source model may have an influence, when uniform source zones are used even substantial differences in the source model cannot match the variability in the hazard produced by the attenuation model. The disaggregations also illustrate perhaps the strongest shortcoming of the Monte Carlo seismic hazard procedure, which is the dominance that the aleatory variability in the attenuation model has on the hazard curve. Models with a higher sigma value clearly tend to increase the dominance of small near-field earthquakes with higher percentiles of ground motion in the hazard analysis. The dominance of small magnitude earthquakes in the disaggregation leaves the user with a problem. The example given for Argostoli, with respect to a ”Cephalonia” earthquake scenario illustrates the inaccuracy in designing for small magnitude earthquakes with high sigma.

Since the aleatory variability in the attenuation cannot be ignored, and assuming uni-

form source zones (or hybrid models), then how can the apparent contradiction of the high hazard originating from smaller events be avoided? One alternative may be to perform a disaggregation using a spectral acceleration at a longer period, possibly greater than 0.15s. Uncertainty in the attenuation is different for each period. At the longest spectral periods fewer records may be available and accelerations from small magnitude events are ill-defined, usually because they are not detected robustly. This may have the impact of increasing the sigma term at long periods, or more likely making the sigma an unrealistic approximation of the true variability in spectral acceleration at longer periods.

A more promising alternative could be to use a duration dependent strong shaking parameter such as Arias Intensity or Cumulative Absolute Velocity. Danciu and Tselentis (2007) have attenuation relations for these parameters, which could be viable given that they are derived from Greek earthquakes. In both instances the functional form remains the same as for PGA, albeit with a raised sigma value. Other Arias Intensity relations with a non-linear magnitude scaling term do exist (Travasarou *et al.*, 2003; Stafford *et al.*, 2008a), but it is not clear that they are appropriate to Greece.

Another alternative could be the use of Intensity as the strong motion parameter. This approach could easily be designed to exclude earthquakes not of engineering significance. Theoretical problems are encountered when treating Intensity as a continuous parameter when it is not diagnosed and recorded as such. The same intensity value may be estimated for different levels of recurrence probability, or a discrete band of recurrence probabilities. Whilst a hazard curve may make the distinction between an intensity of 9.25 and 9.75, differences on this scale are not clearly diagnosed. The problem emerges when considering the tolerance bounds in hazard level being input into disaggregation. Selection of all earthquakes producing an intensity of either VIII or IX for example will span an enormous distance and magnitude range, thus negating the arguments for using it in favour of PGA or other engineering-related parameters.

Disaggregation is possible using a stochastic approach, but the uncertainties in the attenuation permeate through the analysis. The Monte Carlo method can identify design earthquakes, but consideration must be given to the inputs in the analysis. Undertaking different disaggregations for different combinations of source and attenuation model is one way of doing this. It may also be prudent to compare these disaggregations with deterministic scenarios based on knowledge of seismotectonics in a region. These analyses may assist in the identification of hazard models that are most appropriate when compared with the seismotectonics of the Aegean.

6.7 Conclusions

This chapter has demonstrated how the Monte Carlo method of seismic hazard analysis can be used for both mapping and site-specific purposes. The outputs of these analyses,

for a given set of inputs, are compatible with user demands of PSHA in terms of probability of a PGA being exceeded, uniform hazard spectra and design earthquake scenarios. Comparison with previous hazard maps for the Aegean region show good agreement in many parts of Greece and the Balkans. There appears to be some discrepancy in western Turkey, which suggests that none of the source models used here adequately capture the real tectonic behaviour of that region.

Some general comments can be made about the PSHA shown here. Firstly, it is clear that the Am05 attenuation produces acceleration that may be unfeasibly high, for both PGA and spectral acceleration. The Am96, DT07, Bm07 and BA07 relations show better agreement and, when translated into intensity, are consistent with the historical records. It is quite apparent that the appearance of the hazard maps is strongly influenced by the choice of source model. Conversely, for site-specific applications it would appear to be the choice of attenuation model that has the greatest influence.

To further the last point, it has become apparent that the aleatory variability term in the attenuation models has an enormous influence on the hazard analysis. The characterisation of uncertainty in strong ground motion models is a problem in seismic hazard analysis that remains a long way from resolution. The comparison of the DT07, Bm07 and BA07 models illustrates how differences in the sources of the hazard being modelled are not always obvious in the hazard output. The Bm07 and BA07 models were derived from different data sets spanning different regions and with vastly different functional forms. Despite this, the hazard maps and hazard curves show remarkable similarity. DT07 is derived from a much smaller data set. As such, it has a smaller aleatory uncertainty term and produces lower accelerations. The differences in the aleatory uncertainty between the DT07 model and Bm07/BA07 may only be a reflection of the limited data set (limited in terms of area, magnitude range and tectonic conditions). Here, the smaller aleatory uncertainty term in the DT07 model reduces the number of small earthquakes contributing disproportionately to hazard at a site. This reduction may only reflect the input data set and not a reduction in the true variability in SGM attenuation.

The comparison of the impact of different attenuation models in the seismic hazard analysis provides an invaluable insight. Of the three models most commonly used here, each is derived from different data and represents a different region: DT07 from Greece exclusively, Bm07 from Europe and the Middle East and BA07 from a worldwide (if slightly California and Taiwan biased) strong motion dataset. All three relations have been developed within a few months of each other. Prior to this, most European seismic hazard analyses had used Am96, whilst some Greek hazard analyses used Sk03 or Margaris *et al.* (2001) exclusively. The hazard curves show that both the Bm07 and BA07 produce similar curves to Am96. Given the different functional forms of each of these relations this is an encouraging result. On this basis, and given the improved fit to strong motion data given in Chapter 5, it is reasonable to suggest that both Bm07 and BA07 should supersede Am96 as the attenuation model of choice for Europe. However, which of these relations

should be preferred? This may depend on the application as BA07 has a wider spectral range, more suitable for engineering purposes. Conversely, Bm07 has a wider magnitude range, which may make it more suitable if hazard from lower magnitude earthquakes is a consideration (for example aftershock hazard or analyses for shorter return periods).

The use of DT07 is a more complicated decision. It clearly predicts lower ground accelerations than the other models, yet it is also derived from a less diverse data set. It has a limited magnitude range, but a wide spectral range. There is obviously a strong argument for the DT07 relation superseding previous Greek relations, but not necessarily for use in place of Bm07 or BA07, especially given that BA07 proved a better ft to the observed strong motion data.

This chapter has also provided the opportunity to compare the impact of different source models in the seismic hazard analysis. The progression of existing source models (PZ1990 and PP2000) to a novel source model (WT2006) to source models derived from a different method (K-means) is particularly enlightening. PP2000 has previously been the standard model for use in the Aegean yet it does not necessarily compare favourably to these new alternatives. WT2006 produces a better χ^2 fit to the observed data. This is, at least partially, due to the use of observed seismicity in constraining synthetic sources in low-moderate seismicity regions. The K-means models, on the other hand, delineate uniform source zones over a similar area to the PP2000 model, also producing a better fit to the observed seismic hazard. The hazard maps produced using the K-means zones, however, contrast most strongly to existing hazard maps. This might suggest that the partitioning of seismicity is appropriate, yet the actual delineation of the zone boundaries is such that some of the zones have spread hazard over too wide an area. The K-means methodology has been shown to produce new source models that are compatible with existing ones. Nevertheless, the process of translating partitions into uniform source zones needs more refinement before it could be used in place of interpretive models of seismotectonics. It is hoped that future research will help further this goal.

The understanding and interpretation of the attenuation models can, and will, continue to be debated ad infinitum. For the end users of a seismic hazard analysis, such consideration may not be of importance. The provision of a suite of different maps and results may not satisfy the demands of the end user. Judgements can be made as to which combination of source, recurrence and SGM attenuation model may be the most realistic, but this oversimplifies the issue of epistemic uncertainty. It does not diminish nor remove epistemic uncertainty. Furthermore the maps shown here make simplifications with regard to the site characteristic, the style-of-faulting and the temporal behaviour of seismicity.

The next chapter will illustrate how further applications of stochastic techniques and new input data sets can help refine these hazard maps to reflect the real variation in hazard across a region. This will include analysis of epistemic uncertainty, incorporation of fault

data, aftershocks and new databases of site characteristics. The ultimate aim in the next chapter is to be able to identify the ”best” map, which conforms to the demands of the hazard analyst. This should lay the ground work for future developments in Monte Carlo based PSHA.

Chapter 7

Epistemic Uncertainty Analysis and Extensions to the Monte Carlo Method for Seismic Hazard in Greece

The previous chapter has shown how different source models and attenuation models can give contrasting results when input into a seismic hazard analysis. This produces a suite of hazard maps or hazard curves, each of which could be considered as characterising seismic hazard across a region or at a site. Such a deconstruction of the hazard analyses arising from different combinations of models can be an interesting and insightful process. For the end-users of these analyses, it may not be sufficient to be presented with a suite of results and be requested to make a judgement as to which map or curve is preferred. It is perhaps more beneficial to identify a "preferred" curve or map and then present the alternative models. Which ever manner is chosen, it is still necessary to identify a "preferred" combination of source, recurrence and attenuation model, preferably with a quantitative assessment of likelihood. In practice, this is achieved by undertaking an analysis of epistemic uncertainty.

To recap, epistemic uncertainty refers to the uncertainty arising from "incomplete data and knowledge regarding the earthquake process" (Anderson and Brune, 1999). It may also be referred to by the more explicit term "model uncertainty" or "reducible uncertainty". This uncertainty arises when multiple models may be considered as applicable to a process or system. It is not a requirement that each model must be equally applicable to the system.

The hazard maps and curves presented in the previous chapter illustrate some of the scale of epistemic uncertainty in hazard analysis in the Aegean. Within this epistemic uncertainty are five source models (seven including the automatic verses manual parameter input on the PP2000 and PZ1990 models) and four attenuation models (six if including

Am96 and Sk03). There are several other places where it could be argued that alternative models could also be used. The first is in the choice of recurrence relation, where the truncated Gutenberg and Richter (1944) model is used at the exclusion of alternatives (justification for this is given in Chapter 2). Another area is the choice of maximum magnitude for each zone. The value given for maximum regional magnitude (M_W 8.6) is based on a general agreement between different methods. For each zone only the Cumulative Moment method is used. Alternative methods could also have been used, although given the dependence on b-value this would result in a propagation of correlated errors through the analysis. It is common in hazard analysis to select, carefully, particular areas where the choice of model is prevalent. When analysing epistemic uncertainty the areas of uncertainty that are not considered, by deliberate choice or otherwise, contribute to the "unknown unknowns". In essence, any analysis of uncertainty can only treat the modelled epistemic uncertainty as a sample of the true epistemic uncertainty.

Comparison of the automated input verses manual input hazard analyses for the PZ1990 and PP2000 models suggests that differences in recurrence relation and maximum magnitude are overwhelmed by differences in the choice of attenuation and source model. This in itself is a disconcerting observation as it could be interpreted that the size of the largest earthquakes and their frequency of occurrence have little bearing on the hazard at a location. This is clearly not a pragmatic assumption for the purposes of engineering design, but as has been shown in the disaggregations it is clear that the attenuation relation, and its characterisation of intrinsic variability, have an enormous influence.

There are two aims to this chapter. The first is to assimilate epistemic uncertainty into the seismic hazard analysis. This will be attempted using a Monte Carlo approach and the more traditional Logic Tree approach. By incorporating different models (appropriately weighted) into the hazard analysis it is hoped that convergence toward a single hazard map will be achieved. The second aim is to incorporate some modifications to the Monte Carlo procedure to understand what impact, if any, they will have on the seismic hazard analysis. These modifications are intended to add an additional layer of sophistication into the hazard analysis. They include site and fault characterisation, hazard from intermediate-depth earthquakes and non-Poissonian seismicity.

7.1 Analysing Epistemic Uncertainty

7.1.1 Logic Tree

The most common approach to analyse epistemic uncertainty is the use of a logic tree (Coppersmith and Youngs, 1986; Reiter, 1990; Kramer, 1996). This is a method of formalising the different models and indicating the spread of their respective hazard values, visualised in Figure 7.1. Each node of the tree represents a part of the hazard analysis where different models are considered, whilst the branches of the tree at each node

represent the models themselves. Weights are assigned to each branch to represent the relative degree of confidence in each model. Each permutation of models is then given an overall confidence score (the term probability is avoided as the weights are usually assigned by judgement). The total sum of confidences is equal to unity. To arrive at a value for the hazard analysis, the mean or median of the weighted branches is taken. There is disagreement regarding which of the mean (and standard deviation) and median (and inter-quartile range) is preferred. It is argued that use of the mean is invalid as the probabilities in the calculation are measures of confidence, and that at very low annual probabilities of ground motion the hazard becomes dominated by the least likely models (Abrahamson and Bommer, 2005). Conversely, Musson (2005) argues that fractile levels (such as median or 84^{th} percentile) are not truly probabilistic as the specific probability of occurrence of the branch contributing to the level of the fractile ground motion is low. This is also qualified by a call for more judicious pruning of the logic tree (i.e. removing very low probability branches).

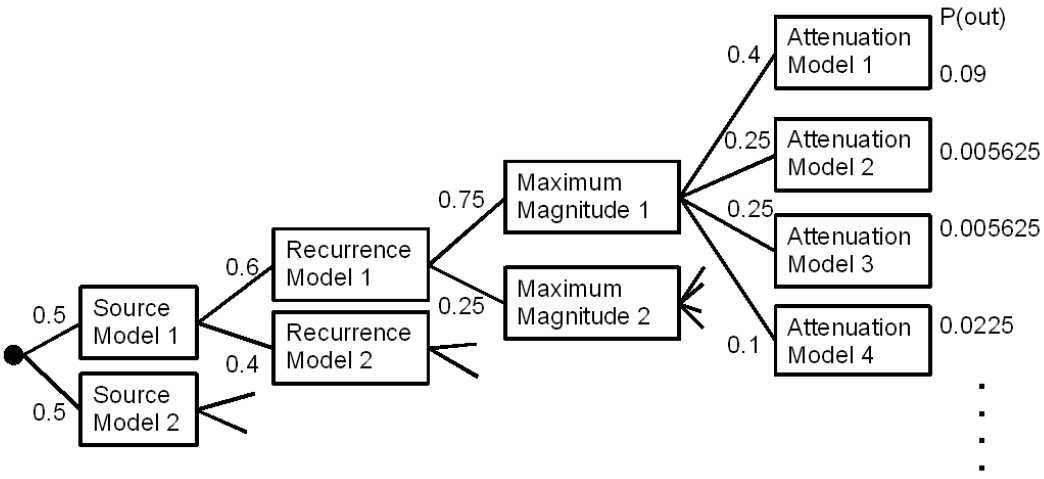

Figure 7.1: Visualisation of a Logic Tree Model.$\sum P\left(out\right) \equiv 1$

7.1.2 Monte Carlo Method 1

An alternative to the logic tree approach is the use of the Monte Carlo techniques. Cramer *et al.* (1996) use Monte Carlo simulation to create a suite of different hazard results, each hazard result created using the Cornell-McGuire method. Each hazard result arises from a particular combination of parameters which are themselves derived from the distribution of uncertainty on the key parameters of the hazard analysis (fault length, b-value etc). The uncertainty is then represented on a map by way of either the $\pm 2\sigma$ range or by the coefficient of variation.

Several aspects of the Cramer *et al.* (1996) method have already been incorporated into the Monte Carlo seismic hazard analyses shown in the previous chapter. These features include uncertainty in b-value, aleatory variability in the attenuation relation, and uncertainty in inter-event time. As with the Cramer *et al.* (1996) method, these uncertainties

were modelled by allowing the parameters to vary according to a normal (or in the case of attenuation, lognormal) random number distribution. This approach is realistic in the sense that the observed uncertainties in the key parameters are also passed into the seismic hazard model. This is extended further by Smith (2003), who uses the Monte Carlo techniques to sample parameters randomly from a frequency distribution to take errors in the parameters into consideration. It is the Smith (2003) approach that is followed more closely in chapter 6, where b-value is sampled from a normal distribution to incorporate uncertainty. It could, however, also be interpreted as mixing sources of epistemic and aleatory variability, thus making the assumption that estimates of b-value and its error, for example, remain constant. This is, perhaps, a little pessimistic as one would expect the error on these parameters to reduce in time. In practice this may still be the most realistic approach to undertaking seismic hazard analysis allowing for the uncertainty in these parameters.

The incorporation of uncertainty on specific parameters via a Monte Carlo method, still does not resolve the issue of which source/attenuation model to use. To achieve this the Monte Carlo technique shall be extended a step further. To understand this, it is necessary to recall why the Monte Carlo technique can be used in probabilistic hazard analysis in the first place. Musson (2005) gives an eloquent description:

> *"If one had the divine ability to see all alternative possible futures of seismicity of the next 50 [T] years, one would observe the actual hazard to the site in terms of ground motions happening more frequently or less frequently. The Monte Carlo hazard process is the next best thing to this vision - an observational method of hazard estimation".*

The obvious proviso to this statement is that the futures of seismicity are not an exhaustive range of scenarios, but a finite range of seismicity scenarios as defined by the inputs to the model. Each T year simulation of seismicity is a realisation of a controlled random process within the bounds of the model set up. This alone does not answer the question as to what the model should look like, only how this translates to hazard estimation. A question that should be asked is whether the input model needs to be fixed for all simulations? This would produce seismic hazard results, as seen in the previous chapter, and even takes into consideration some epistemic uncertainty by allowing for variation of parameters within each model. If different models are considered as being possibly applicable [but not necessarily equally applicable] to a region then you simply arrive at a suite of hazard analyses with very little guidance as to which is preferred. This was produced in the last chapter. If, on the other hand, the assumption that each realisation of seismicity must conform to the same model is relaxed, the Monte Carlo technique can be used to integrate different models of seismicity.

7.1.3 Monte Carlo Method for Analysis of Epistemic Uncertainty (MCMAEU)

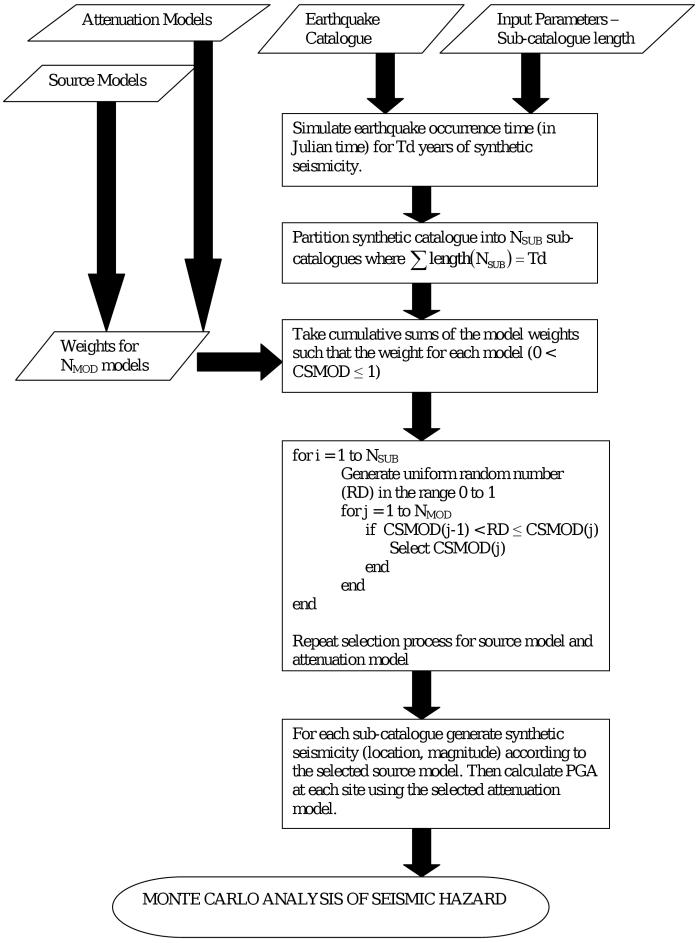

Figure 7.2: Flowchart of the MCMAEU procedure

To allow the input models to vary, the synthetic catalogue must be partitioned into N_{SUB} sub-catalogues (essentially the realisations of seismicity described in Musson (2005)). As described in the previous chapter, the synthetic time distributions of the catalogues in this approach are not dependent on the input model, but instead the input catalogue, which remains constant. This essentially anchors seismicity rate to the observed regional distribution. The relative number of earthquakes in each zone within the synthetic time is then used to define the proportion of events belonging to each zone. This method of creating a synthetic time distribution means that the errors in a-value for all the zones in the model do not accumulate to produce wildly fluctuating annual rates of seismicity. It also means that the time distribution of the whole synthetic catalogue can be created irrespective of the source model. A synthetic catalogue can therefore be partitioned into sub-catalogues prior to the assignment of location and magnitude, which are dependent on the selected model.

To select the model for each sub-catalogue a weighted random selection method is used,

similar to those commonly found in genetic algorithms (Coley, 2005). Since the cumulative sum of the model weights must be unity, the model weight is re-assigned on the basis cumulative weight rather than absolute weight. A roulette selection is then used whereby a uniform random number is generated in the range 0 to 1. The first model for which the cumulative weight is greater than the random number is then selected as the model. This process is illustrated in the flowchart in Figure 7.2. The selection process is repeated for source model and attenuation model for all the sub-catalogues.

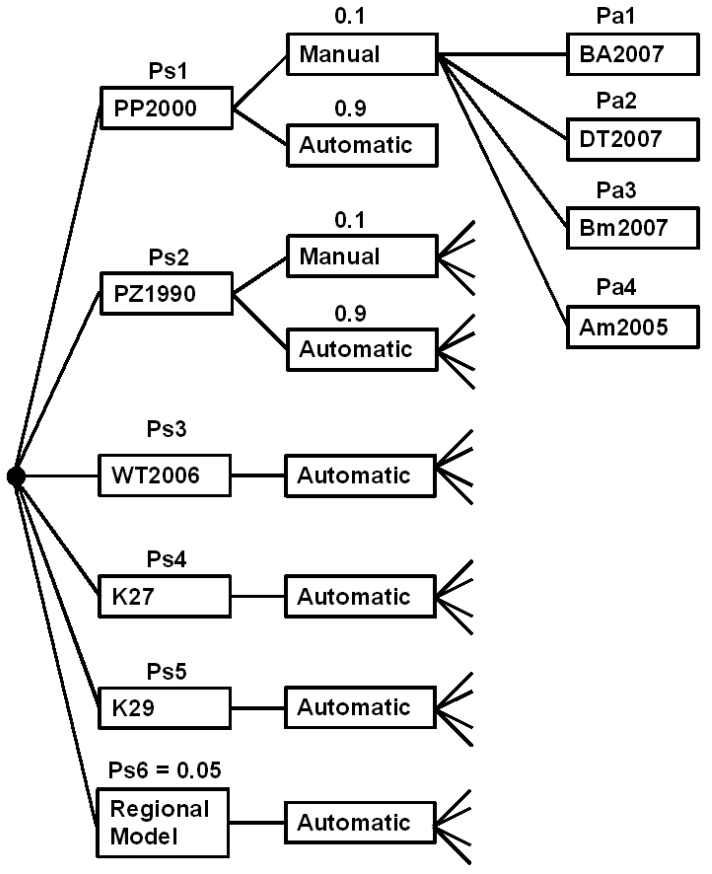

Figure 7.3: Logic Tree schematic of the epistemic uncertainty analysis performed in this chapter.

One advantage of the Monte Carlo method proposed here is that it can be visualised in a logic tree format (shown here in Figure 7.3). Essentially each sub-catalogue is selecting a specific path along the logic tree. Providing that a large number of sub-catalogues are used, and that the whole synthetic catalogue is of sufficient duration as to allow for a reasonable length of sub-catalogue (in this example greater than 50 years), then this method should effectively sample the range of source and attenuation model combinations. The hazard value that is output from the Monte Carlo analysis takes into account epistemic uncertainty. This method also avoids making decisions regarding the choice of mean or median hazard, as would be the case if using the logic tree.

Essentially, this method is combining the Monte Carlo approach to epistemic uncertainty modelling with the formalism of model confidences associated with a logic tree. The

Table 7.1: Source model weightings used in this analysis

Source Model	χ^2	Weight (Based on χ^2)	Weight (SOCRATES)
PP2000 (automatic)	1.368	0.06	0.1
PZ1990 (automatic)	1.264	0.12	0.1
WT2006	1.157	0.3	0.2
K27	1.294	0.2	0.2
K29	1.282	0.2	0.2
PP2000 (manual)	1.605	0.04	0.1
PZ1990 (manual)	1.360	0.08	0.1

output is, again entirely compatible with the demands of PSHA, as the hazard calculation is performed in the same way as it is if one model is used.

7.1.4 Model Weights

There are three areas of model uncertainty within the seismic hazard analysis: 1) the choice between regional seismicity and the source zones, 2) the choice of source model and 3) the attenuation model. The choice between zoned and regional seismicity has already been implemented in the previous analysis by randomly selecting 5 % of the events in the model and then re-sampling the earthquakes from the zone-free regional distribution of seismicity. This remains the same in the present models.

For the attenuation model, four different weighting schemes were presented in Chapter 5. These were based on the likelihood fit of the attenuation to observed strong motion data from the Aegean region. The four weighting schemes represented are 1) EXPONENTIAL [EXP], 2) EXPONENTIAL-PLUS [EXP+], 3) LINEAR [LIN] and 4) SOCRATES [SOC]. The SOC weighting accounts for the situation where no preference for any model is assumed; hence all the weights are equal.

To determine which, if any, of the source zone models should be given higher weightings the values of χ^2 fit to observed seismicity are used as a guide. These χ^2 were determined using the method described in Chapter 4 and have already been shown in Chapter 6. A lower χ^2 value is associated with a higher weighting, or a similar weighting to other models if the differences are very small. These weightings are shown in Table 7.1. As discussed previously, the WT2006 model produces the lowest χ^2. Some difficulties arise in this method. The two K-means derived models have lower χ^2 values than the PZ1990 model if the manually input parameters are used, but higher values than for the automatically derived parameters. The K-means models are therefore given equal weightings to the PZ1990 model, but a stronger weighting is given to the automatically derived parameter PZ1990 model than the manually input parameter model. The PP2000 model gives higher χ^2 values than all the others and is therefore given the lowest weighting. Once again the automatically derived model parameters give lower χ^2 values than the manually input ones, which is reflected in the weighting scheme.

A second weighting scheme is also considered alongside the χ^2 weighting scheme. This second scheme assumes that all models are equally weighted, and is once again referred to as the SOCRATES weighting scheme [SOC]. Comparison of the two weighting schemes shows there is actually very little difference between the two schemes, with the WT2006 afforded a stronger weighting and PP2000 a weaker weighting in the χ^2 scheme. This illustrates the need for this epistemic uncertainty analysis since there is little justification to remove any of the source models suggested. It is likely, in fact, that several more K-means models could be added to this epistemic uncertainty analysis.

The weighting schemes presented here are all based on quantitative assessment of the fit of source and attenuation models to observed seismicity and strong motion, respectively, in the Aegean region. The decision was made to avoid a complete "Expert Judgement" assignment for these weighting schemes, since there is little guidance about how expert judgement should be apportioned. This does not mean that expert judgement cannot be input into the process. More weighting schemes where weights have been assigned by judgement could also be appended on to these schemes *ad infinitum*. Instead, focus shall only be placed upon the two source model weighting schemes and four attenuation model schemes.

7.1.5 Seismic Hazard using Monte Carlo Analysis of Epistemic Uncertainty

As with the single model analyses shown in Chapter 6, PGA hazard maps are generated for the 10 % probability of being exceeded in 50 years. The same maps for 5 % and 2 % probability of being exceeded in 50 years are found in Appendix B. Four different approaches are used:

1. Comparison of the χ^2 (weighting based on χ^2 fit to observed hazard) and SOC source model schemes using the BA07 attenuation model (Figure 7.4).
2. Comparison of the EXP, EXP+, LIN and SOC attenuation schemes using the PP2000 source model (Figure 7.5)
3. Comparison of the EXP, EXP+, LIN and SOC attenuation schemes using a variable source model with χ^2 weighting (Figure 7.6).
4. Comparison of the EXP, EXP+, LIN and SOC attenuation schemes using a variable source model with SOC weighting (Figure 7.7).

The faulting regime is assumed to be normal or normal/strike-slip, depending on how it is defined in the respective attenuation relation. The site is considered to be a "rock" site (NEHRP Class B, V_{s30} 760 $m\ s^{-1}$). Horizontal ground acceleration is assumed to be geometric mean of the two components; hence Am05 and Bm07 have been adjusted down accordingly.

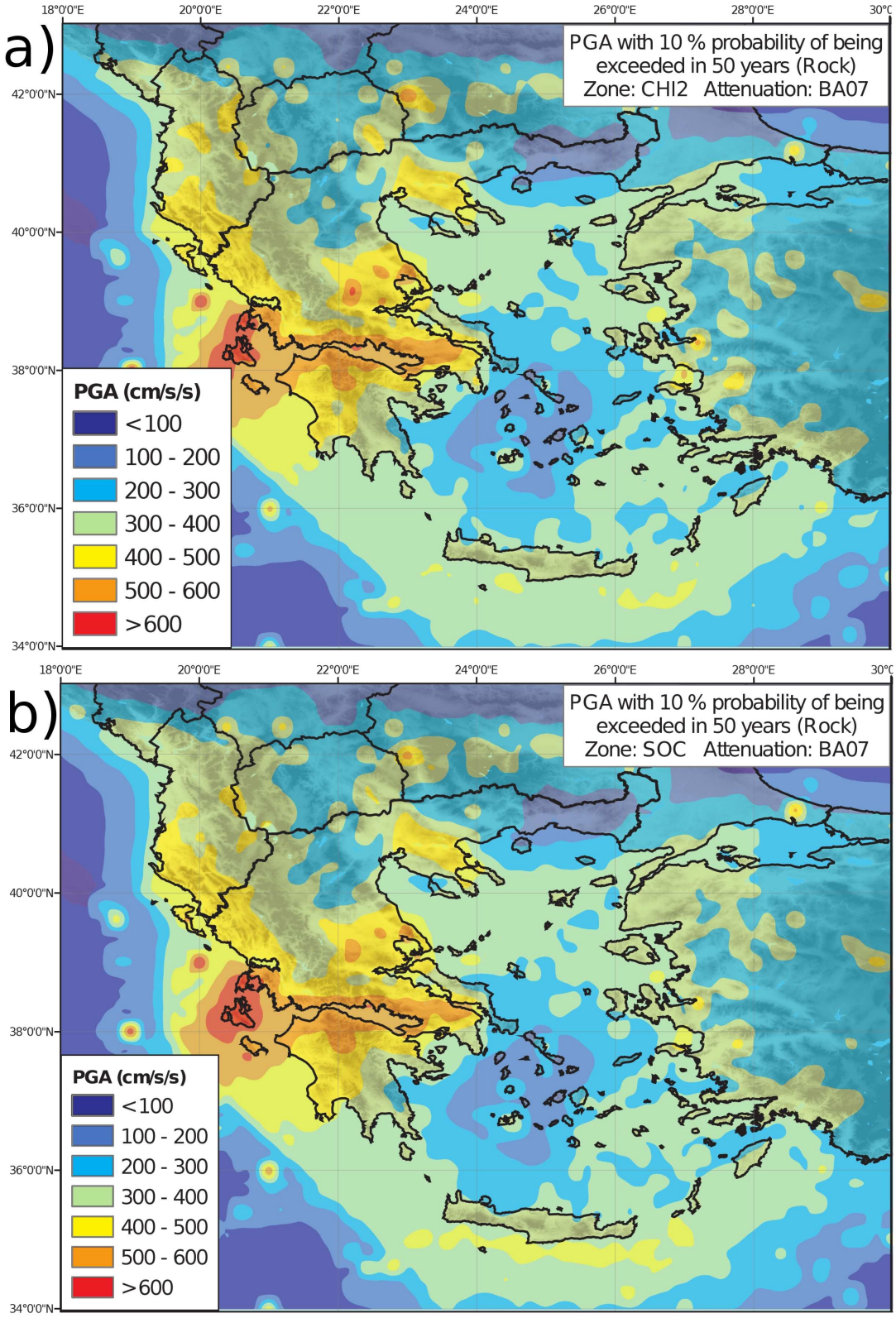

Figure 7.4: MCMAEU calculation of PGA with a 10 % probability of being exceeded in 50 years for an assumed rock site and normal/strike-slip faulting using the BA07 attenuation model. The χ^2 zone weighting scheme is used in a)and the SOC weighting scheme in b)

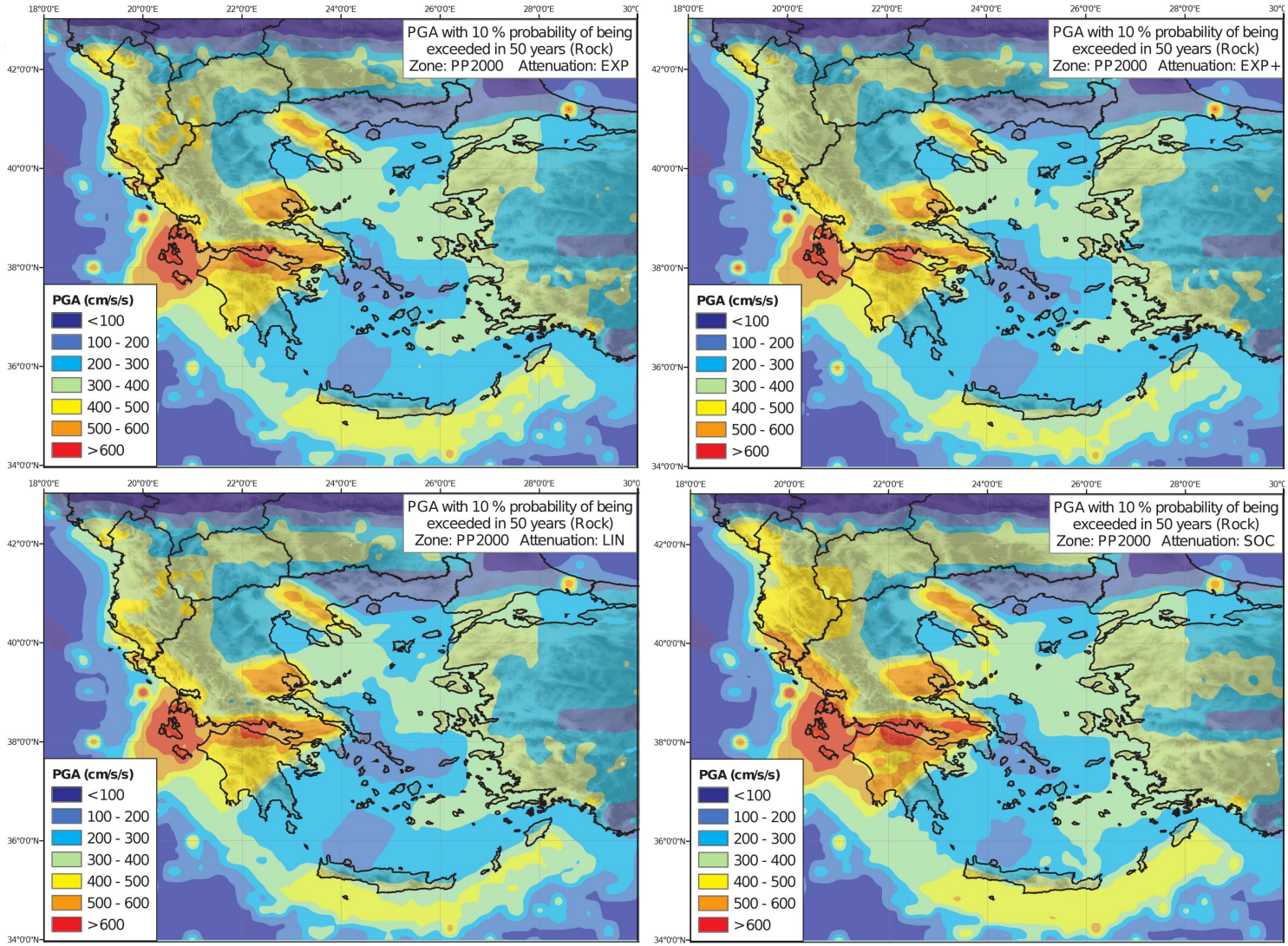

Figure 7.5: PGA with a 10 % probability of being exceeded in 50 years (rock site, normal/strike-slip faulting, PP2000 source model). Attenuation weighting schemes are EXP (top left), EXP+ (top right), LIN (bottom left) and SOC (bottom right).

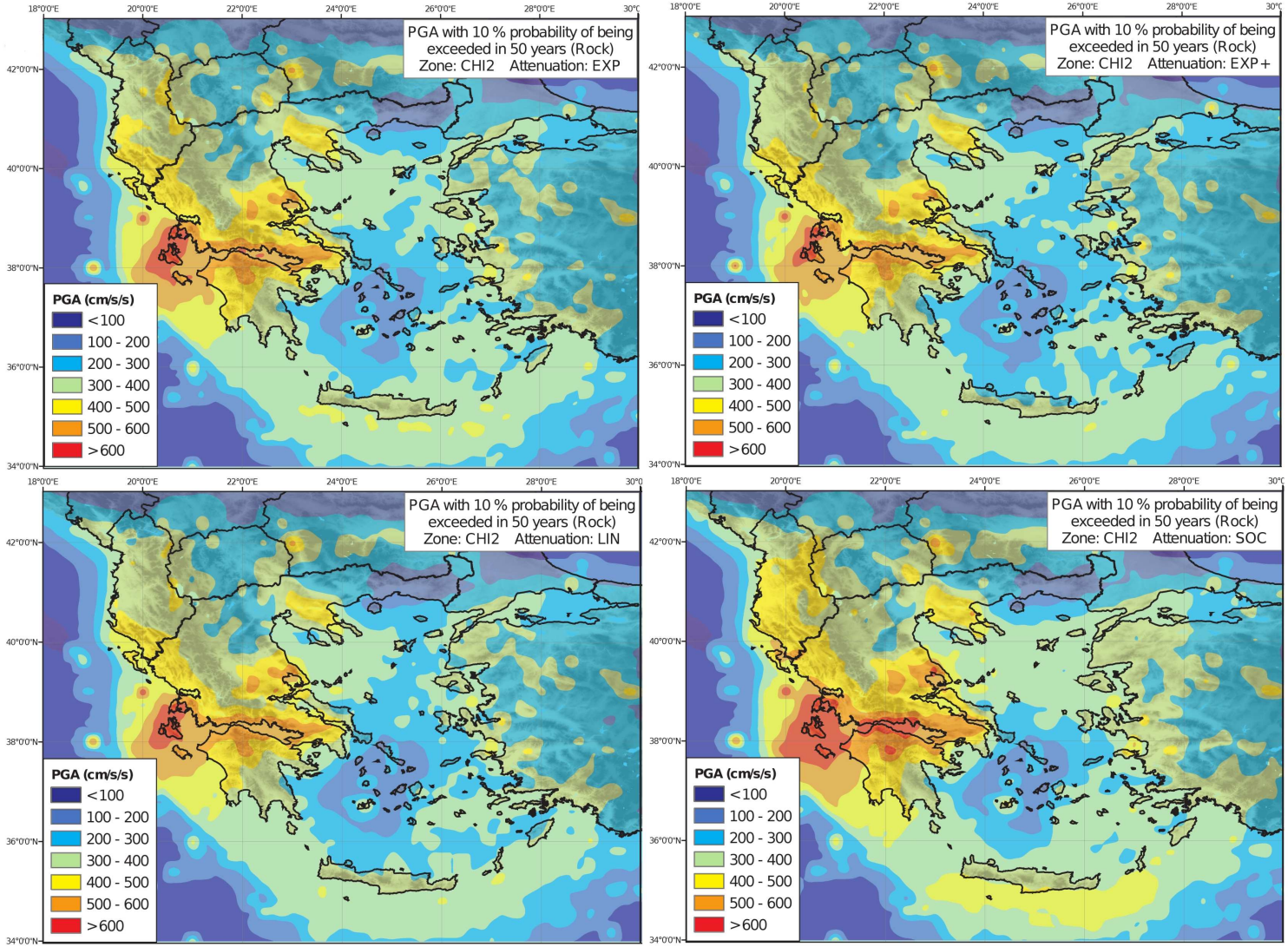

Figure 7.6: PGA with a 10 % probability of being exceeded in 50 years (rock site, normal/strike-slip faulting, χ^2 source model weighting scheme). Attenuation weighting schemes are EXP (top left), EXP+ (top right), LIN (bottom left) and SOC (bottom right)

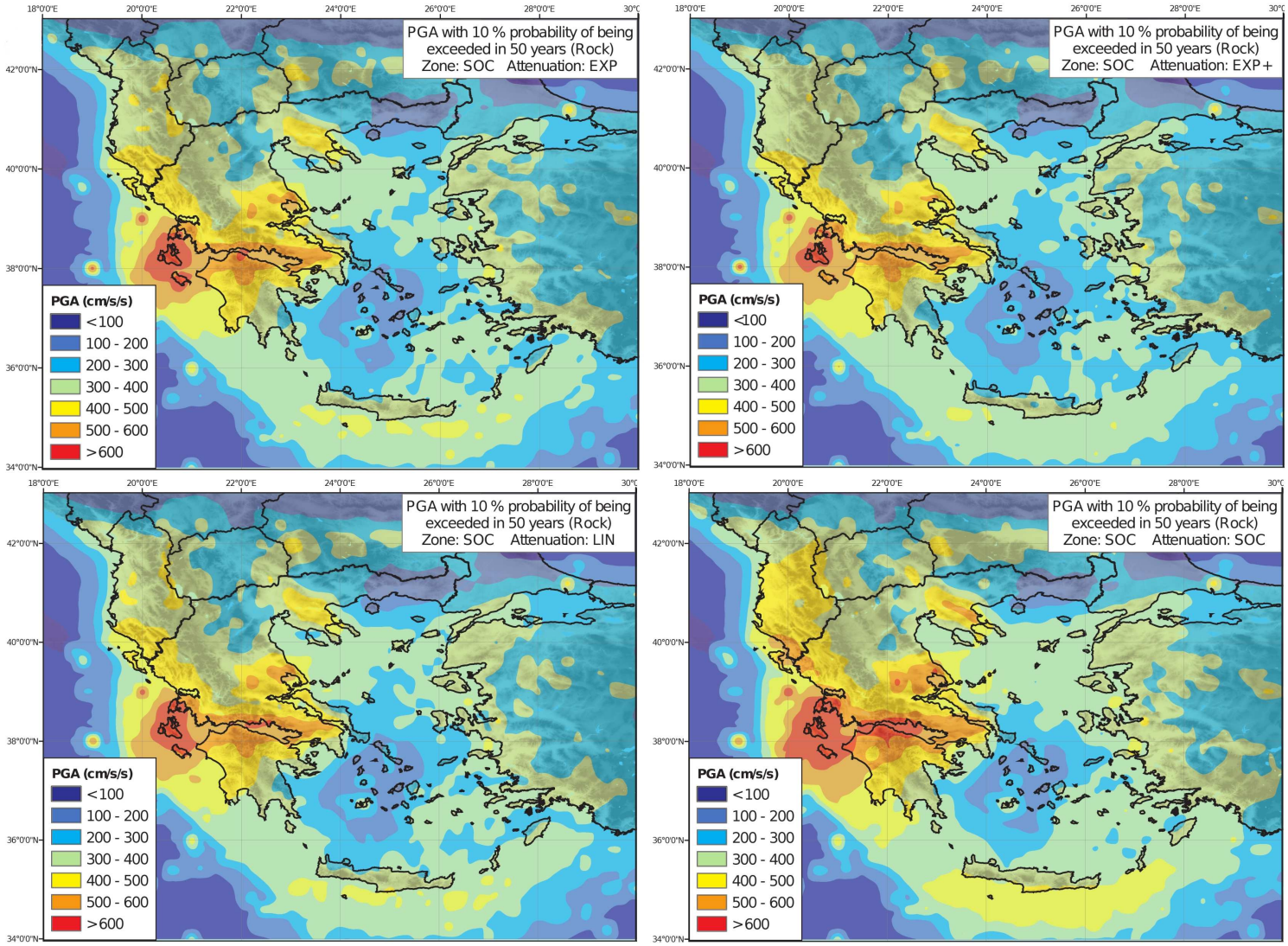

Figure 7.7: PGA with a 10 % probability of being exceeded in 50 years (rock site, normal/strike-slip faulting, SOC source model weighting scheme). Attenuation weighting schemes are EXP (top left), EXP+ (top right), LIN (bottom left) and SOC (bottom right)

The most obvious feature of the maps in Figures 7.4 to 7.7 is how little difference small variations in the weightings make to the hazard. If Figures 7.4a and 7.4b are compared then there is very little difference between the hazard maps when either χ^2 or SOC zone model weighting scheme is used. This in itself is not surprising since the two weighting schemes are similar for most of the zones. The similarities would suggest a robustness of the hazard analysis to small variations in the weighting scheme. When maps produced with variable models are compared with those using a fixed model, the obvious effect is that the regular shape of the zones is less apparent in the hazard map. In the PP2000 hazard maps there are several areas of raised seismic hazard that broadly correspond to the shape of particularly active zones. These regular shapes are much less obvious in the zone-variable maps. In effect, by combining different zone shapes into one map, the zones are essentially being softened by virtue of variation in shape. This would suggest, by fortunate accident, that another way to soften zone boundaries in PSHA would be to average across different models, thus preserving the contrast in seismic behaviour but avoiding the blocky appearance due to hard boundaries.

Some differences are more obvious when the four attenuation weighting schemes are applied. Again there are considerable similarities in the maps, but the different attenuation weighting schemes do have an observable impact. It would appear that the EXP and SOC weighting schemes tend to result in higher levels of ground acceleration for much of the Aegean than do the EXP+ and LIN schemes. This effect is more obvious in the maps of 5 % and 2 % probability of being exceeded in 50 years. It is particularly pronounced in the Ionian and Gulf of Corinth regions, but is also visible around much of the plate margin (Adriatic coast round to the eastern Hellenic arc).

To understand why these differences occur due to the attenuation weighting scheme it is necessary to recall the observations made for the single models in the previous chapter. Allowing for conversion from larger horizontal component PGA to geometric mean PGA, there were several consistent trends that emerged. The Am05 relation predicted considerably higher levels of PGA than the other relations. Conversely, DT07 produced consistently lower levels of PGA for most of the sites considered. The BA07 and Bm07 relations produced very similar hazard curves and were intermediate values in the range considered. From this it can be deduced that a higher weighting for the Am05 relation will pull the hazard values towards a higher PGA, whilst the DT07 relation will pull them towards lower PGA. The SOC weighting scheme gives Am05 equal weighting to the other relations. This scheme tends to produce the higher levels of PGA, especially for longer return periods, which is a direct effect of the influence of the Am05 relation. In the EXP+ scheme the effect of the Am05 relation is negligible (0.02). Here the BA07 relation is dominant (with a weighting of 0.7) although the DT07 relation has enough weighting (0.2) to produce a slight decrease in the PGA; hence the slightly lower PGA on the hazard maps. The LIN scheme produces lower PGA for the same reason, i.e. that the stronger weighting of the DT07 relation (0.3) more than offsets the slight increase in the Am05 weighting (0.1), thus tending towards lower PGA. Finally the EXP scheme

produces very similar maps to that of the EXP+ and LIN schemes, albeit slightly higher. Again this is due to the slightly stronger influence of the Am05 relation offset against the slight weakening of the DT07 relation.

The differences in the attenuation weighting schemes appear to be more pronounced at lower probability levels. This is most obvious in the SOC weighting scheme maps. The reason for this is that the PGA values calculated for the longer return periods at a site are increasingly likely to arise from the extreme models. In this case the Am05 could be considered the low probability, high acceleration, extreme model. In the Monte Carlo procedure the accelerations are sorted in descending order of strength, with the lowest probabilities attributed to the highest ranked accelerations. This has two effects, firstly there is the greater variability in the PGA (the noise on the hazard curves) at these probabilities, second is the increasing influence of extreme events. In these simulations an extreme event might be a moderate to large earthquake in the near-field, sampled with a high sigma value in the attenuation relation. If these come from the Am05 model then the PGA will be higher. This pushes same type of events further down the ranking if a different attenuation model is used. What is interesting about this facet of the procedure is that this effect is common to both the Monte Carlo and Cornell methods of PSHA. Since the influence of the more extreme and unlikely models is greater for very low probabilities, then it is necessary to question whether these models should be included at all. In the case of the Am05 model, which does not have a weight in excess of 0.1 in all but the SOC scheme, this is a reasonable step to take.

7.1.6 Hazard Curves

As with the hazard maps, the same Monte Carlo analysis of epistemic uncertainty can be used to produce PGA hazard curves. A single curve is produced for each weighting scheme; hence the impact of different schemes on hazard at a site can be compared. The curves for the 8 sites used in Chapter 6 are shown in Figures 7.8 to 7.10. For three source model scenarios (χ^2, SOCRATES and PP2000) five curves are compared. These correspond to each of the four attenuation model weighting schemes, plus a single curve for the BA07 relation. Finally, a comparison of just the zone weighting schemes is presented in Figure 7.11, using only the BA07 attenuation model.

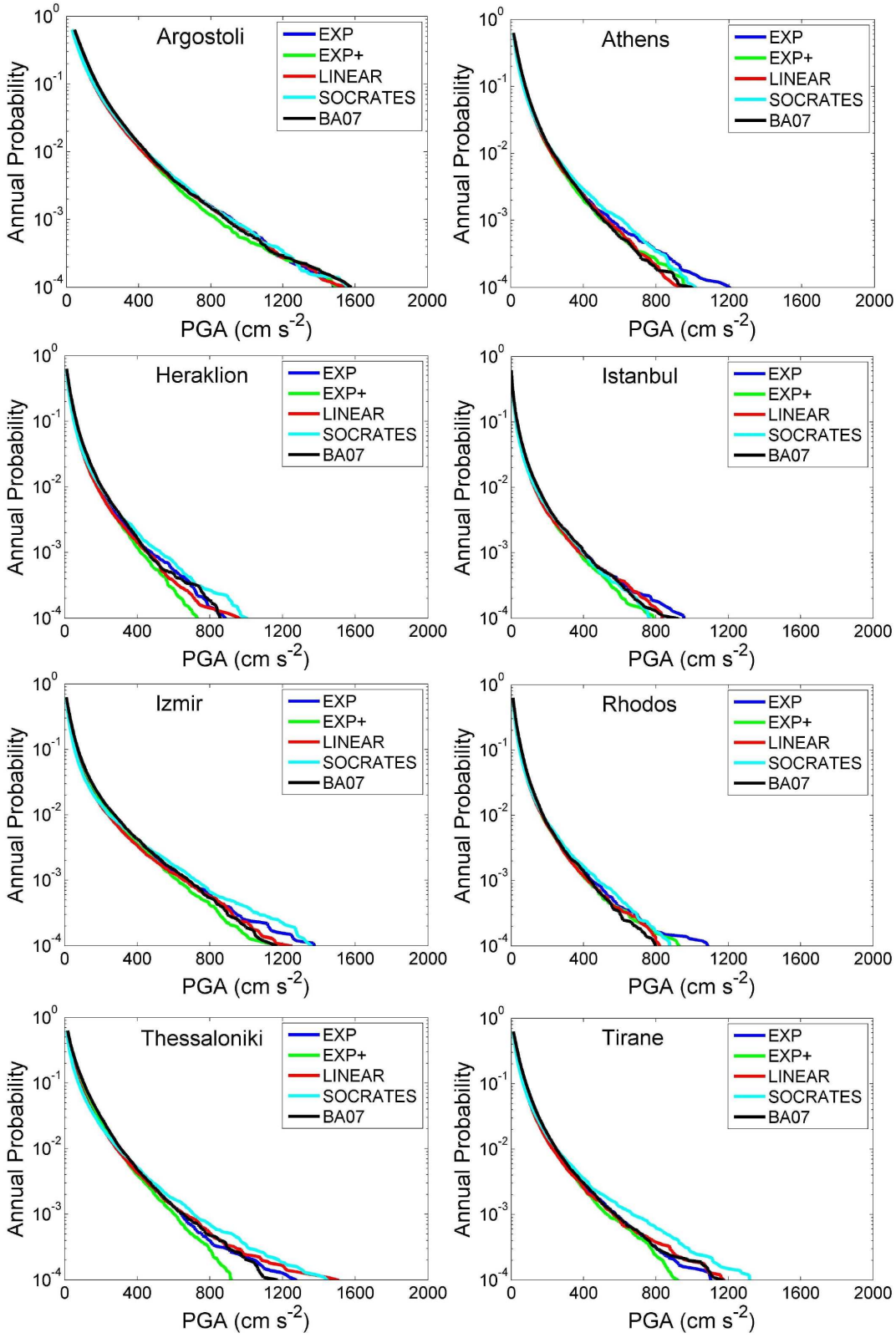

Figure 7.8: Comparison of hazard curves using different attenuation model weighting schemes. χ^2 zone model weighting scheme used, and normal/strike-slip faulting assumed.

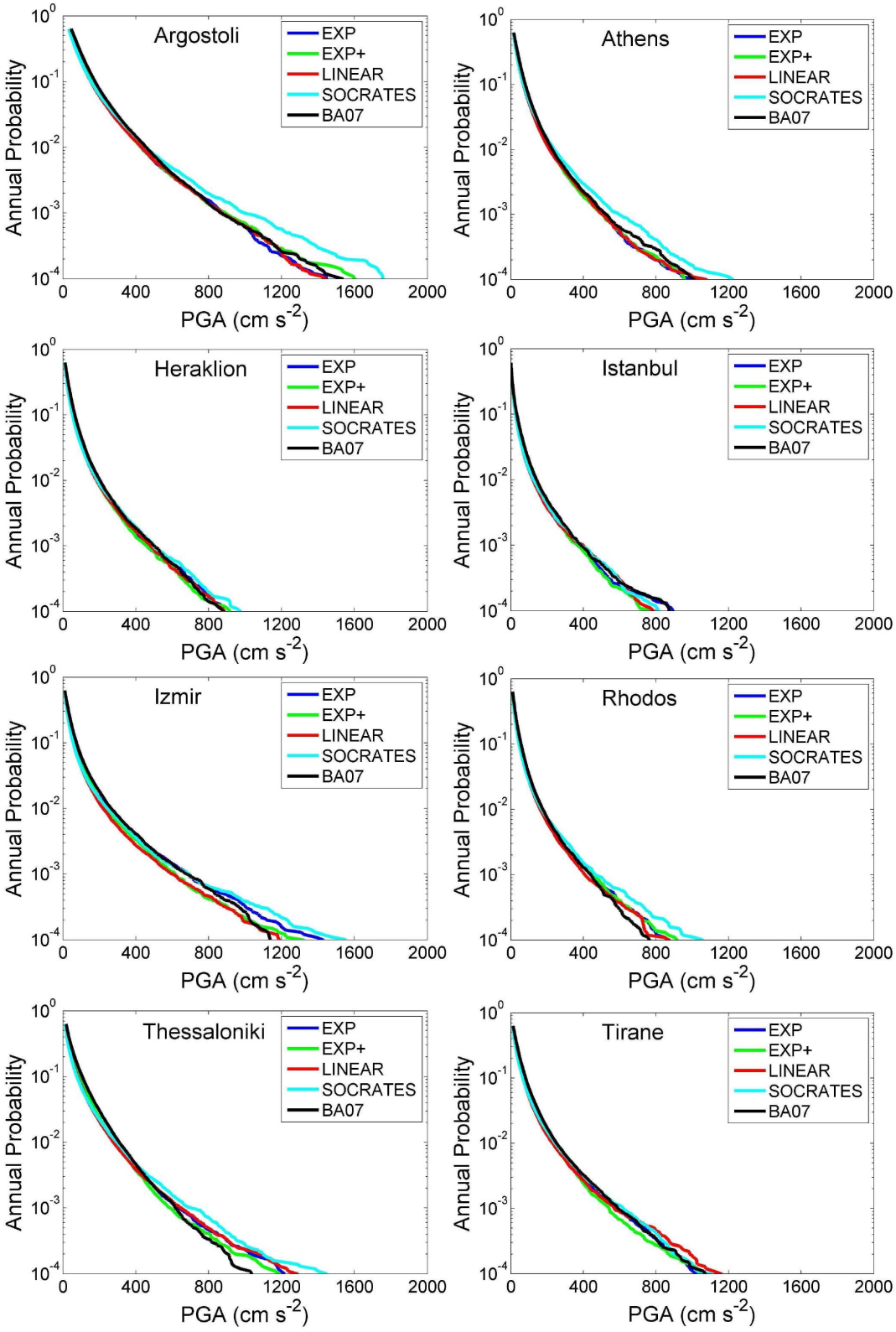

Figure 7.9: As Figure 7.8, using the SOC zone model weighting scheme.

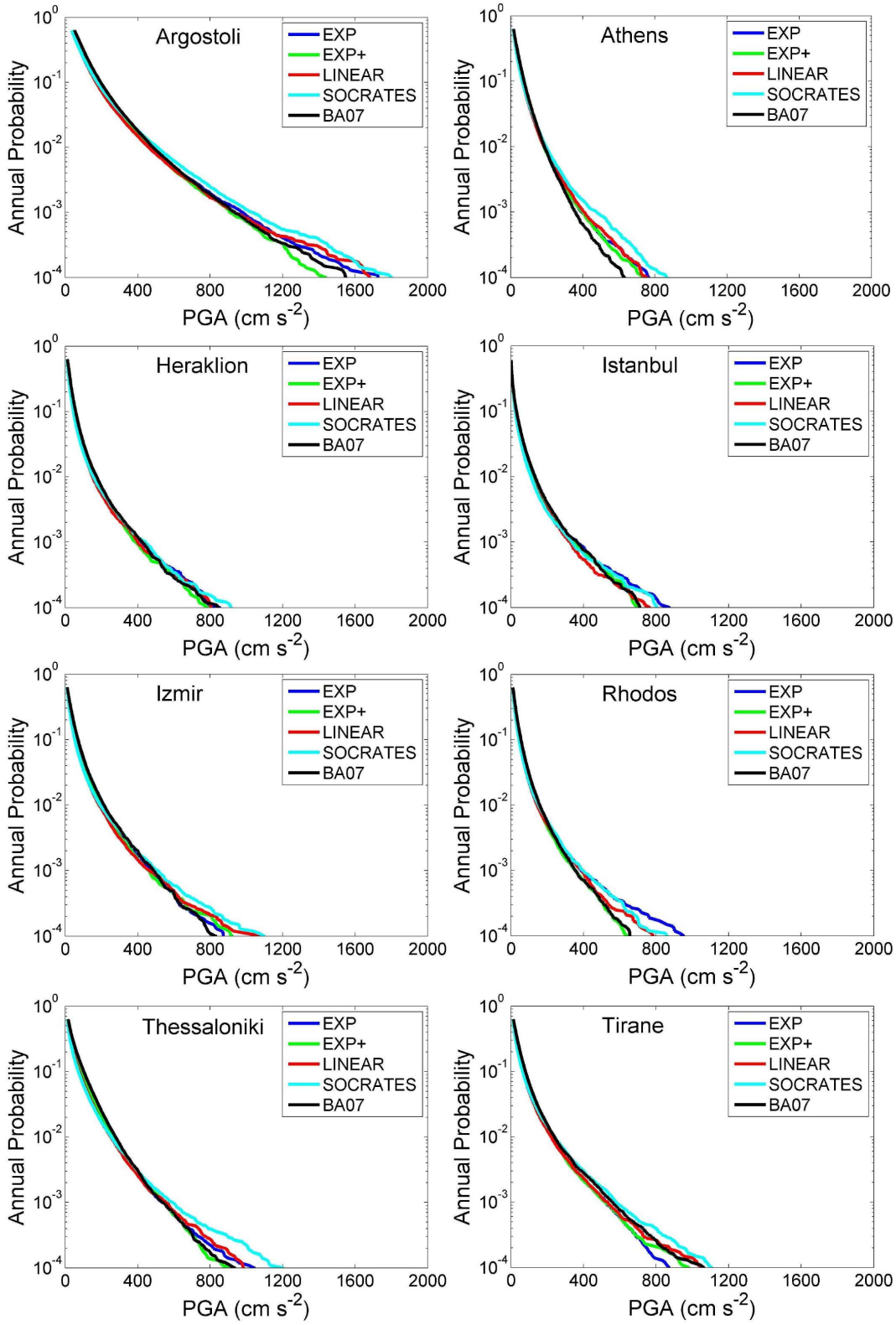

Figure 7.10: As Figure 7.8, using only the PP2000 zone model.

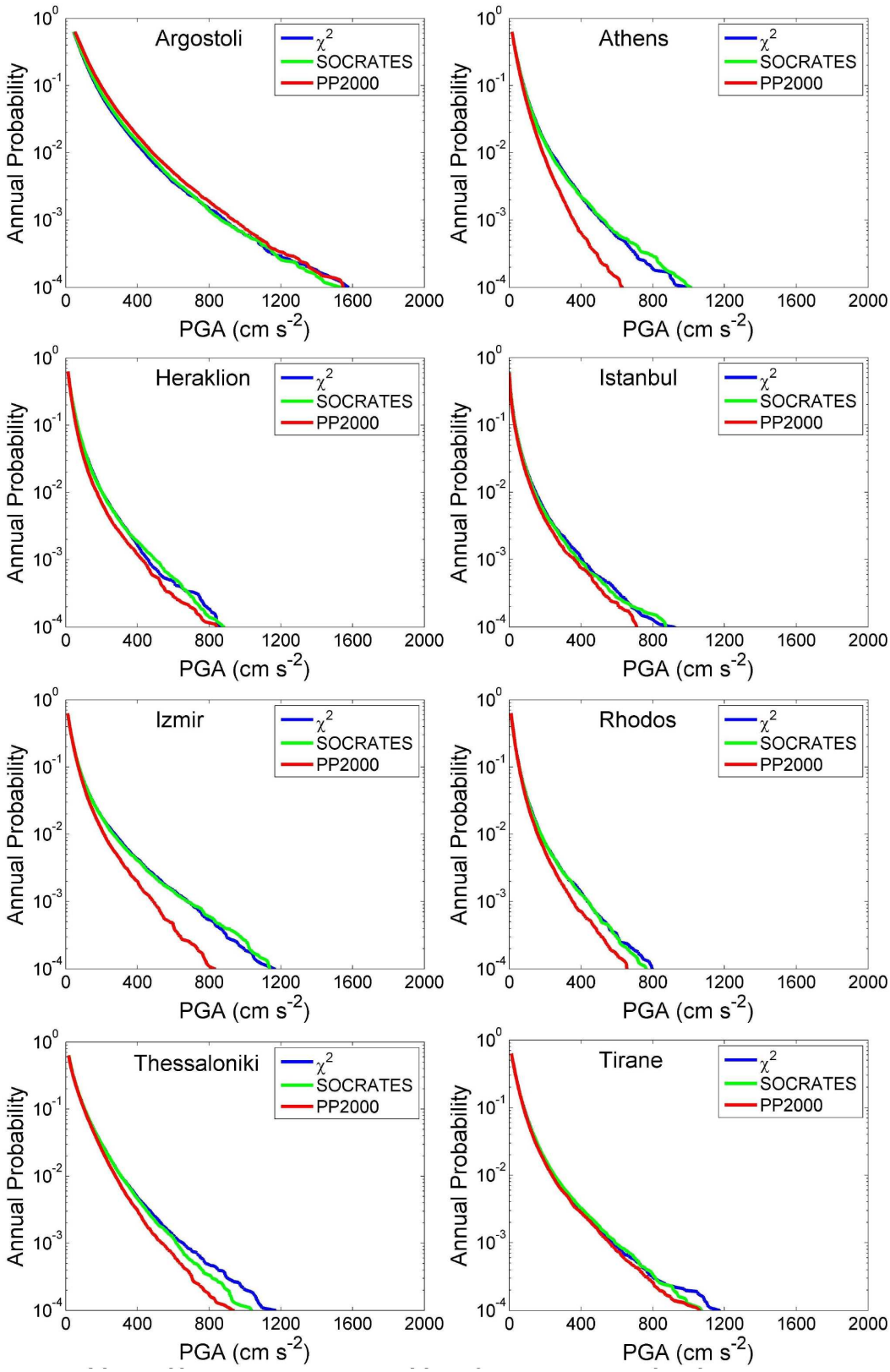

Figure 7.11: Comparison of Hazard Curves when for different zone model weighting schemes. BA07 attenuation model is used and normal/strike-slip faulting assumed.

Once again, these hazard curves reinforce the trends that were suggested in the hazard maps. For a specific zone weighting scheme, there is reasonable agreement between the five curves for probabilities higher than 10^{-3}. Divergence between the attenuation weighting schemes is much more pronounced at lower probabilities. There does not appear to be any robust trend as to which of the weighting schemes produces higher or lower PGA for a site at longer return periods. In most cases the SOC or the EXP scheme tend to produce higher accelerations for probabilities lower than 10^{-3}, whilst the EXP+ scheme produces the lowest. This generally bears out the trends suggested by the maps, but it does demonstrate how variable this trend is from region to region.

The comparison of the zone models shown in Figure 7.11 demonstrates the strong similarity in hazard curve if either the χ^2 or the SOC zone weighting schemes are used. For most sites the curves are near identical, the only divergence being apparent at low annual probabilities at Thessaloniki. The only significant difference between the two weighting schemes is in the comparative weight given to the WT2006 and PP2000 models. In the WT2006 model Thessaloniki sits on the corner of a rectangular shaped uniform zone of high seismic activity around the Thessaloniki-Rentina Fault System. In the PP2000 model it is on the boundary between a large zone of low activity (Kozani zone) and a smaller zone of high activity (Volvi zone). Since WT2006 is given a stronger weighting in the χ^2 scheme then the PGA at low probabilities is higher. Although similar model variations are found elsewhere in the Aegean, the contrast is not as great as it is for Thessaloniki.

7.1.7 Comparison with logic tree methods

Before making judgements regarding the effectiveness of the Monte Carlo method in modelling epistemic uncertainty, comparison is needed with the logic tree methodology. Assuming that the weightings for each model are the same in the logic tree and the Monte Carlo analysis, there is still disparity regarding the harvesting of seismic hazard from the logic tree. As discussed in section 7.1, there is disagreement within the seismological community regarding the use of mean or median hazard of model results in the logic tree. Recognising that there are merits to both approaches, and consequently withholding judgement as to which is preferred, the Monte Carlo method of uncertainty analysis is compared with both the weighted mean hazard and the median hazard. The Logic Tree schematic of the χ^2 zone model and EXP attenuation model combination is shown in Figure 7.12.

The data used in the maps in Figure 6.4 form the basis for the input into the Logic Tree analysis. As before, PGA determined from the Bm07 and Am05 models are scaled down for compatibility with geometric mean horizontal component of PGA attenuation. Three different hazard maps, using the same weighting scheme, are shown in Figure 7.13, whilst maps of the difference in PGA between these methods are shown in Figure 7.14.

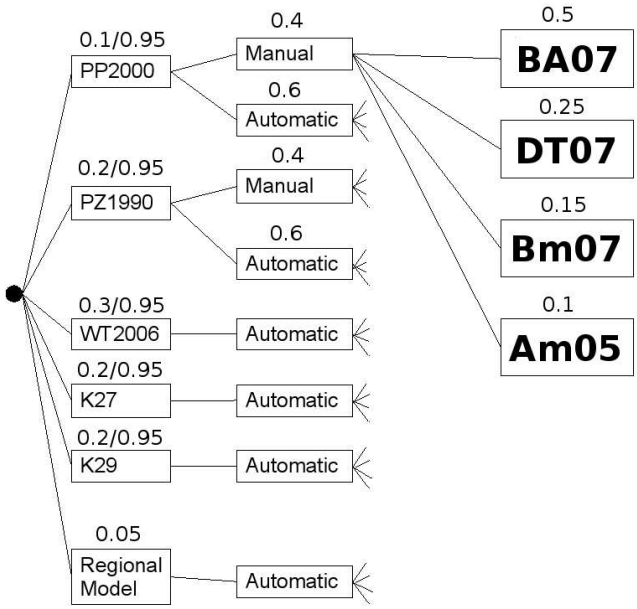

Figure 7.12: Logic tree depiction of the χ^2 zone, EXP attenuation uncertainty model

Upon initial inspection of the hazard maps, there does not appear to be a significant difference in the PGA with a 10 % probability of being exceeded in 50 years, regardless of whether it is Monte Carlo, mean hazard or median hazard that is used. Under the contouring scheme defined here the maps are very similar. The most notable difference between maps is that the logic tree based maps are generally smoother, with fewer islands of high or low hazard than the MCMAEU hazard map. This is because the MCMAEU method uses one synthetic catalogue for the calculation of hazard at each grid point. In the Logic Tree methods the hazard at each site is considered separately, hence the mean or median hazard at one point may correspond to a different synthetic catalogue. This observation would suggest that for MCMAEU analysis, the synthetic catalogues need to be longer in duration than each of the catalogues used in single model analysis.

The maps showing the spatial difference in hazard when the mean, median or MCMAEU methods are used reveal more detail than the hazard maps. For all three maps there is no robust trend as to the location of great disparities between the different techniques. The only areas where mean hazard is substantially and uniformly higher than median hazard, and MCMAEU is higher than both mean and median, are in southern Serbia and the Ionian Sea west of the Ionian islands. It is not clear why these areas should stand out in particular. The only feature common to both of these regions is that they sit outside the source zones delineated by any of the models.

The maps in Figure 7.14 also reveal the differences between the methods. Although there is spatial variability in the difference in hazard for the grid points when different methods are used, the following trends are observed. Mean hazard is, on average, slightly higher than median hazard (mean difference is 8 cm s^{-2}, modal bin is 0 - 10 cm s^{-2}). MCMAEU

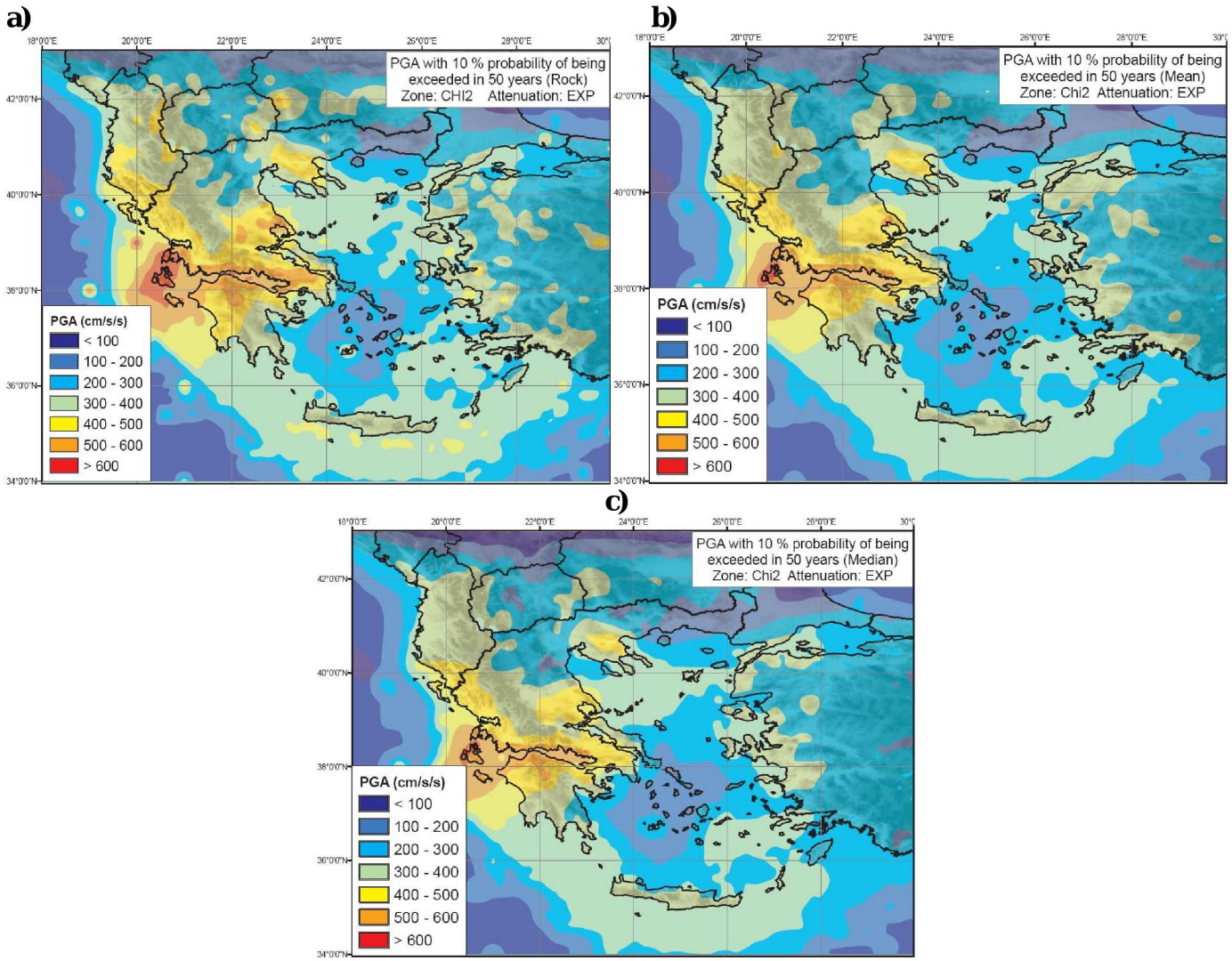

Figure 7.13: PGA with a 10 % probability of being exceeded in 50 years (χ^2 zone model weighting, EXP attenuation model weighting, rock site, normal/strike-slip faulting). a) MCMAEU method, b) logic tree mean and c) logic tree median.

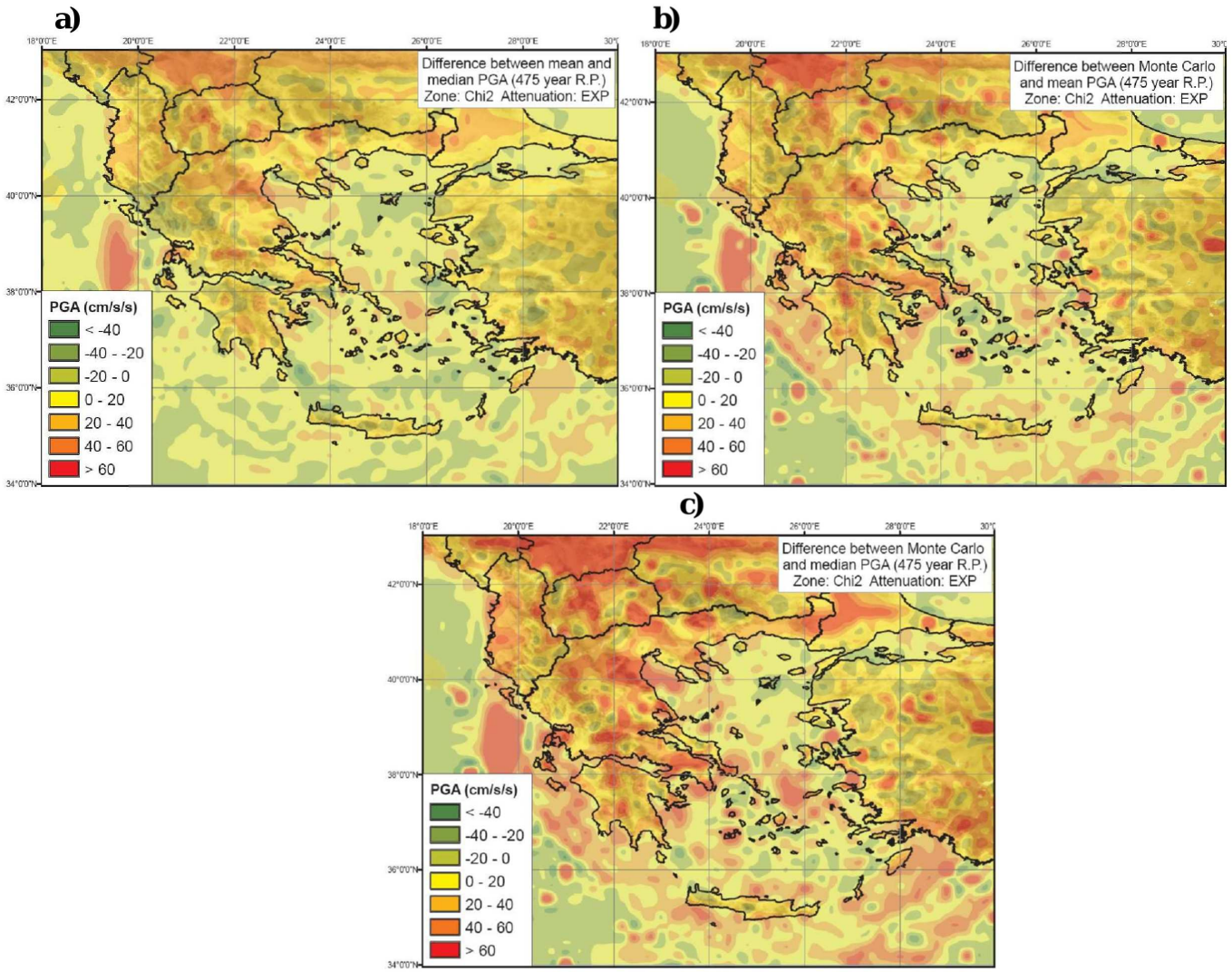

Figure 7.14: Differences in hazard for different epistemic uncertainty measures: a) logic tree mean logic tree median, b) MCMAEU logic tree Mean and c) MCMAEU logic tree median.

is higher than both mean hazard (mean difference of 13.5 $cm\ s^{-2}$, and modal bin 0 - 10 $cm\ s^{-2}$) and median hazard (mean difference of 21.6 $cm\ s^{-2}$ and modal bin of 10 - 20 $cm\ s^{-2}$). There is generally better agreement between the MCMAEU method and the logic tree mean rather than median. This is due to the influence of the Am05 attenuation relation in the mean hazard calculations, whereas this low probability high hazard model will have less influence on the median hazard.

These maps illustrate that MCMAEU can give hazard analysis results that are comparable to, if slightly higher then, logic tree methods of epistemic uncertainty analysis. Two questions emerge. The first question is whether the MCMAEU method is compatible with the demands of seismic hazard analysis as set out in building codes? The second question is whether the conservatism, albeit small, of the MCMAEU method is justified. The application of MCMAEU to PGA hazard shown here provides results that are compatible with the single model Monte Carlo seismic hazard analysis shown in the previous chapter. In that respect, design earthquakes can also be identified using the disaggregation method presented in chapter 6. These can be used to produce design spectra if required. It should be noted, that within the disaggregation, particular magnitude-distance pairs are decoupled from specific source models. For compatibility with building codes it will be necessary for the source zone and attenuation models to be recorded for each earthquake contributing to the specific level of hazard at a site. This may help identify or eliminate models that are not well suited to hazard analysis at a site. It may be more prudent, however, to draw separate disaggragation plots for different source and attenuation model combinations, even if the disaggregation is performed from one synthetic catalogue.

What is currently missing from the MCMAEU approach is the ability to develop a uniform hazard spectrum that incorporates the epistemic uncertainty in the attenuation model. The absence of a UHS is due to limitations of the ability to scale between spectral relations using larger component horizontal spectral acceleration and those using geometric mean (or an alternative) horizontal spectral acceleration. The scaling factor increases with spectral period, i.e. larger component accelerations are larger than geometric mean horizontal accelerations at longer periods. This is compounded by the differences in the spectral ordinates, and the spectral range, used in each attenuation relation. These problems can be overcome by identifying the scaling factor required for a defined set of spectral ordinates. By interpolating all the spectra onto these reference ordinates, and then applying the scaling, the MCMAEU approach can be used to generate a UHS. This is not implemented here for two reasons. Firstly, although the PGA scaling is given in Bommer *et al.* (2005), a reference set of spectral ordinates, with corresponding scaling factors, is not identified. Secondly, the inclusion of the Bm07 relation significantly limits the applicable spectral range. Extrapolation of the attenuation relation beyond the spectral limits defined in Bm07 results in an artificial bias. This can be seen in the shape of the UHS shown for the Bm07 relation in Figure 6.12. In a different region, or with different attenuation relations, and with known scaling factors for predetermined ordinates, a

UHS can be developed using the MCMAEU approach. The inclusion of a UHS should therefore make the MCMAEU approach compatible with the demands of seismic hazard analysis as expressed in many building codes.

The question of whether the conservatism of the MCMAEU method is appropriate, when compared to logic tree methods, is harder to answer. It must be recognised that whilst logic trees are useful, a decision needs to be made as to whether it is the mean, median or a higher fractile of ground motion that is used for design purposes. Such a decision is not made in the MCMAEU method since the epistemic uncertainty is incorporated into the random simulation. Consequently the hazard level that is output by the MCMAEU method is a probabilistic figure that is arrived at by Monte Carlo simulation across the range of appropriate source models. It is, in effect, integrating across the range of magnitude, distance and attenuation variability, as well as a discrete distribution of models. That the resulting hazard levels should be higher than the mean or median is entirely expected as more uncertainty has been integrated into this method; suggesting the conservatism demonstrated in the MCMAEU method is appropriate. This comes at the risk of losing transparency in the hazard analysis. It is important in application of this method to make the end users aware of the hazard analyses for each of the single model results, in addition to the analysis after inclusion of epistemic uncertainty.

7.2 Sensitivity to Minimum Magnitude

7.2.1 Hazard Maps

For the seismic hazard analyses presented so far, a minimum magnitude of M_W 5.2 has been assumed. The reasons for selection of this magnitude level were three-fold: 1) level of magnitude completeness for the instrumental (1900 - 2005) catalogue, 2) Threshold of damage to class A (adobe) structures, defined using macroseismic intensity and cumulative absolute velocity; 3) lower magnitude limit of most of the attenuation relations used here, except DT07 (M_W 4.8) and Bm07 (M_W 3). More detail is found in chapters 2 and 5. Whilst there is good cause for selecting this magnitude level, it is of interest to assess the impact that the choice of M_{MIN} has on the seismic hazard analysis (Bender and Campbell, 1989; Grünthal and Wahlström, 2001; Beauval and Scotti, 2004).

To do this a single source model (PP2000) is selected and two analyses compared (7.15): M_{MIN} = M_W 4.8 and M_{MIN} = M_W 5.2. This is a modest magnitude range to consider, but it spans the likely range of earthquakes magnitude that may be of interest in seismic hazard analysis, without extrapolating attenuations too far beyond their applicable limits. Nevertheless, a decrease in M_{MIN} from M_W 5.2 to M_W 4.8 results in a substantial increase in the number of earthquakes simulated in the synthetic catalogue.

From the seismic hazard maps shown there are very few differences in hazard when

M_{MIN} drops from M_W 5.2 (Figure 7.15a) to M_W 4.8 (Figure 7.15b). What differences there are appear to be spatially variable (Figure 7.16). One interesting observation is that reducing M_{MIN} appears to produce a greater increase in hazard in low-moderate seismicity regions, and slight decrease in the region of highest activity. Given the relatively short return-period being considered in these maps (475 years) the increase in hazard in some regions is to be expected because of the increased contribution of small earthquakes to the hazard at a site. This is consistent with the observations of Grünthal and Wahlström (2001) and Beauval and Scotti (2004).

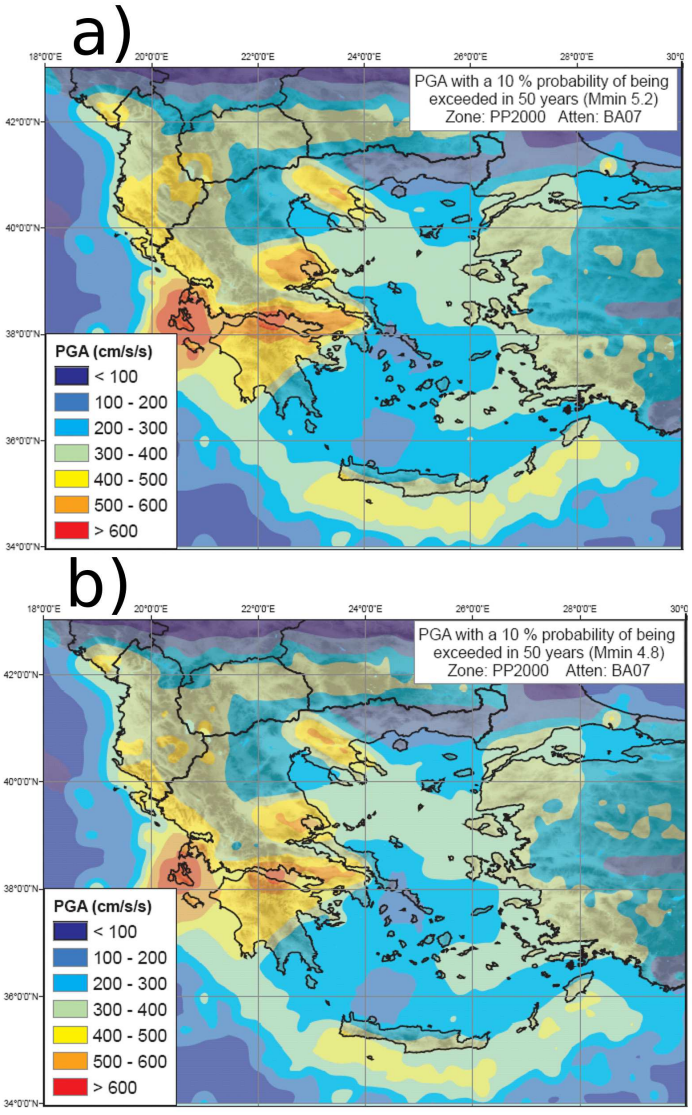

Figure 7.15: PGA with a 10 % probability of being exceeded in 50 years (PP2000 zone model, BA07 attenuation model, rock site, normal/strike-slip faulting). a) M_{MIN} = M_W 5.2, b) M_{MIN} = M_W 4.8

Exactly why the hazard is reduced in southern Greece and the Ionian island is less clear. It is noticeable that b-values are generally higher in this region than in western Turkey. This will increase the contribution of small earthquakes to hazard at a site. Given the fixed aleatory variability term in the attenuation relation, it is possible that the earthquakes contributing to the hazard at that probability level are smaller in the M_{MIN} 4.8

hazard analysis than in the M_{MIN} 5.2 analysis, thus producing smaller ground accelerations even at 2σ to 3σ above the median. The fact that the difference in hazard in southern Greece is not quite as extreme when the Bm07 relation is used might suggest the attenuation relation is an important factor when considering M_{MIN}.

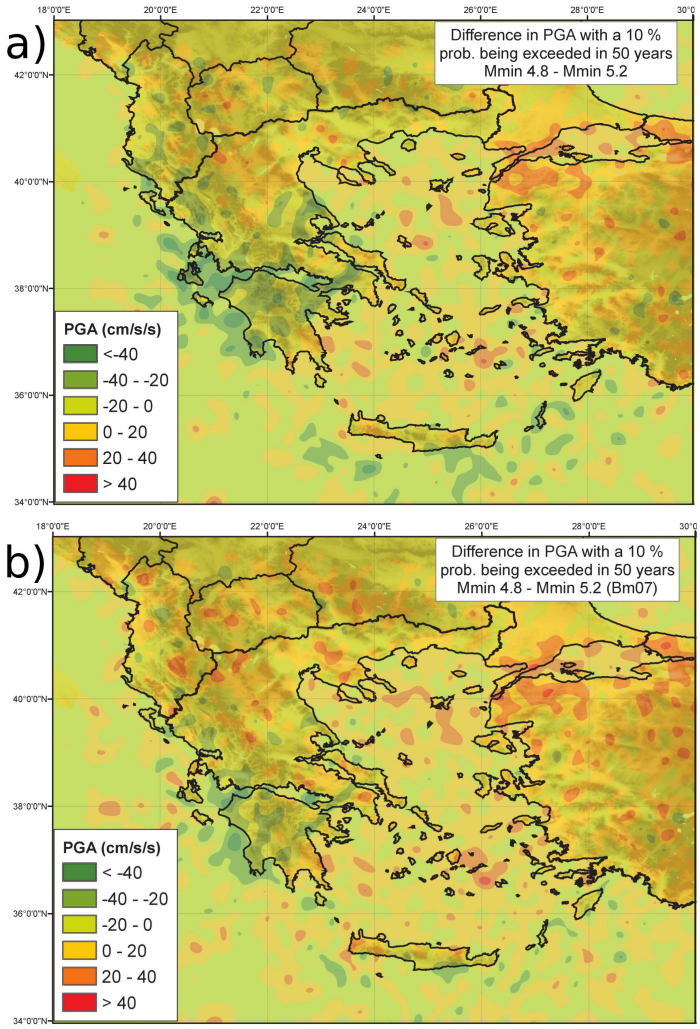

Figure 7.16: Change in PGA (with a 10 % probability of being exceeded in 50 years) when M_{MIN} is decreased from M_W 5.2 to M_W 4.8. a) BA07, b) Bm07

7.2.2 Hazard at a site

One important consideration when comparing this study on the impact of M_{MIN} with those of Bender and Campbell (1989), Grünthal and Wahlström (2001) and Beauval and Scotti (2004) is that each of these studies was implemented in a low-moderate seismicity region (eastern U.S., Germany and France respectively). Their observation of increased hazard with smaller M_{MIN} is generally borne out in the lower seismicity areas of the Aegean considered here. However, to observe the impact of M_{MIN} on the hazard curve and UHS, six sites from across the Aegean are compared. The hazard curves and UHS for each site are shown in Figure 7.17 and 7.18, respectively. The hazard curves in Figure

7.17 are plotted on dual logarithmic axes to elucidate the difference in hazard at shorter return periods.

The hazard curves all display the same trend that a lower M_{MIN} increases hazard at higher annual probabilities (return periods less than 100 years). At longer return periods curves converge and, at some sites, the higher M_{MIN} produces a greater hazard. The exact probability at which the curves merge does appear to vary from site to site. Whether this point is greater or less than the 475 year return period used in the maps is clearly dependent on site.

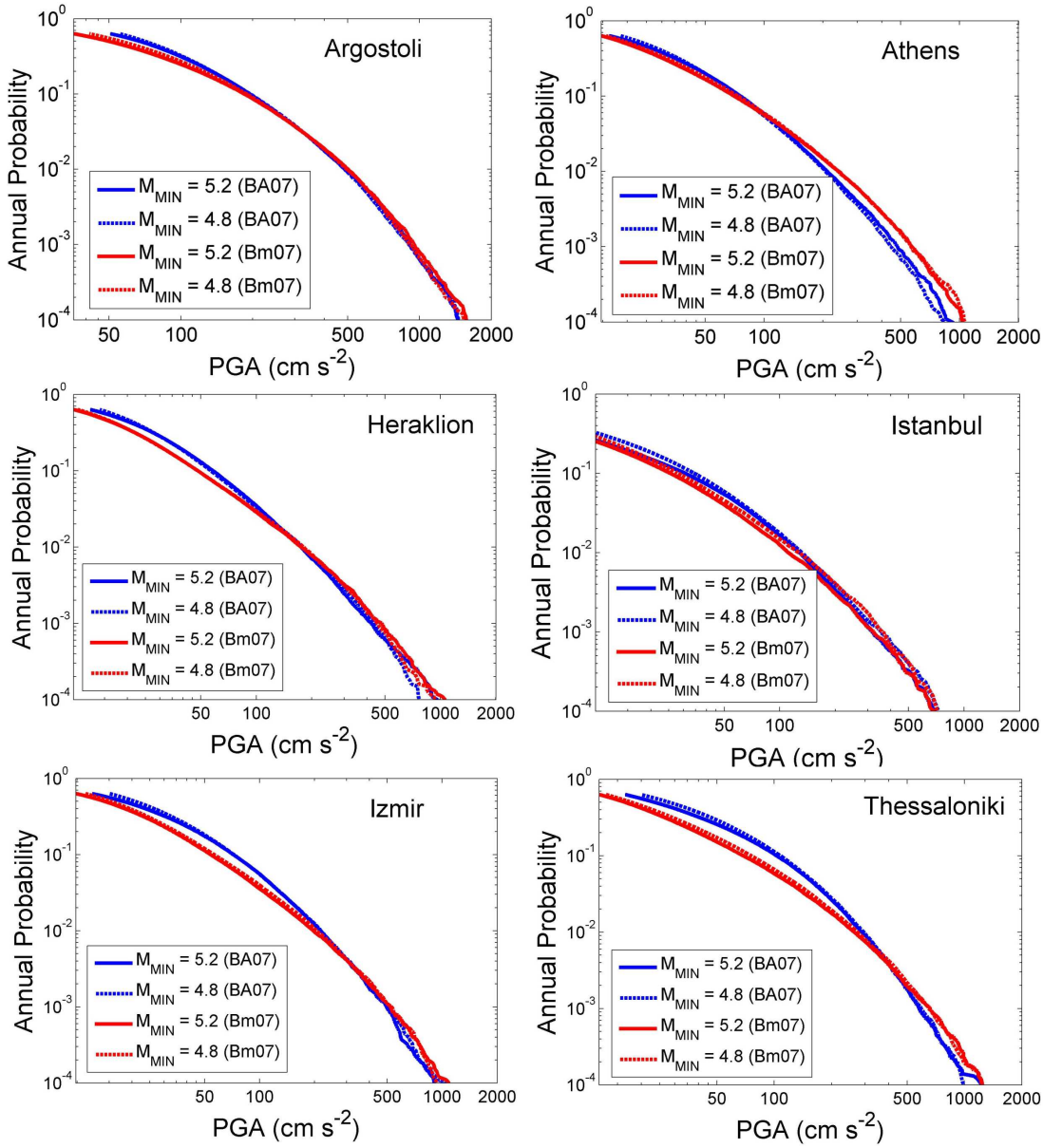

Figure 7.17: Comparison of hazard curves for M_{MIN} = M_W 5.2 (solid line), and M_{MIN} = M_W 4.8 (dashed line). Curves for the BA07 relation are shown in blue and Bm07 in red. PP2000 zone model and normal/strike-slip faulting assumed.

The impact of M_{MIN} on the UHS is not quite as obvious. The Bm07 and BA07 relations

are both shown here. The UHS show considerable variability. The only discernable trend is that for the lower M_{MIN} the UHS is usually slightly higher at periods of less than 0.15 - 0.2 s. This is easily explained by the dominance of higher frequency accelerations from small earthquakes when observed at short distances. Above 0.2 s the hazard spectra are higher for higher M_{MIN}, with the BA07 curve showing convergence at very long periods. The exact difference in spectral acceleration, and the period at which the UHS for the higher M_{MIN} exceeds that for the lower M_{MIN}, clearly depends on the site being considered.

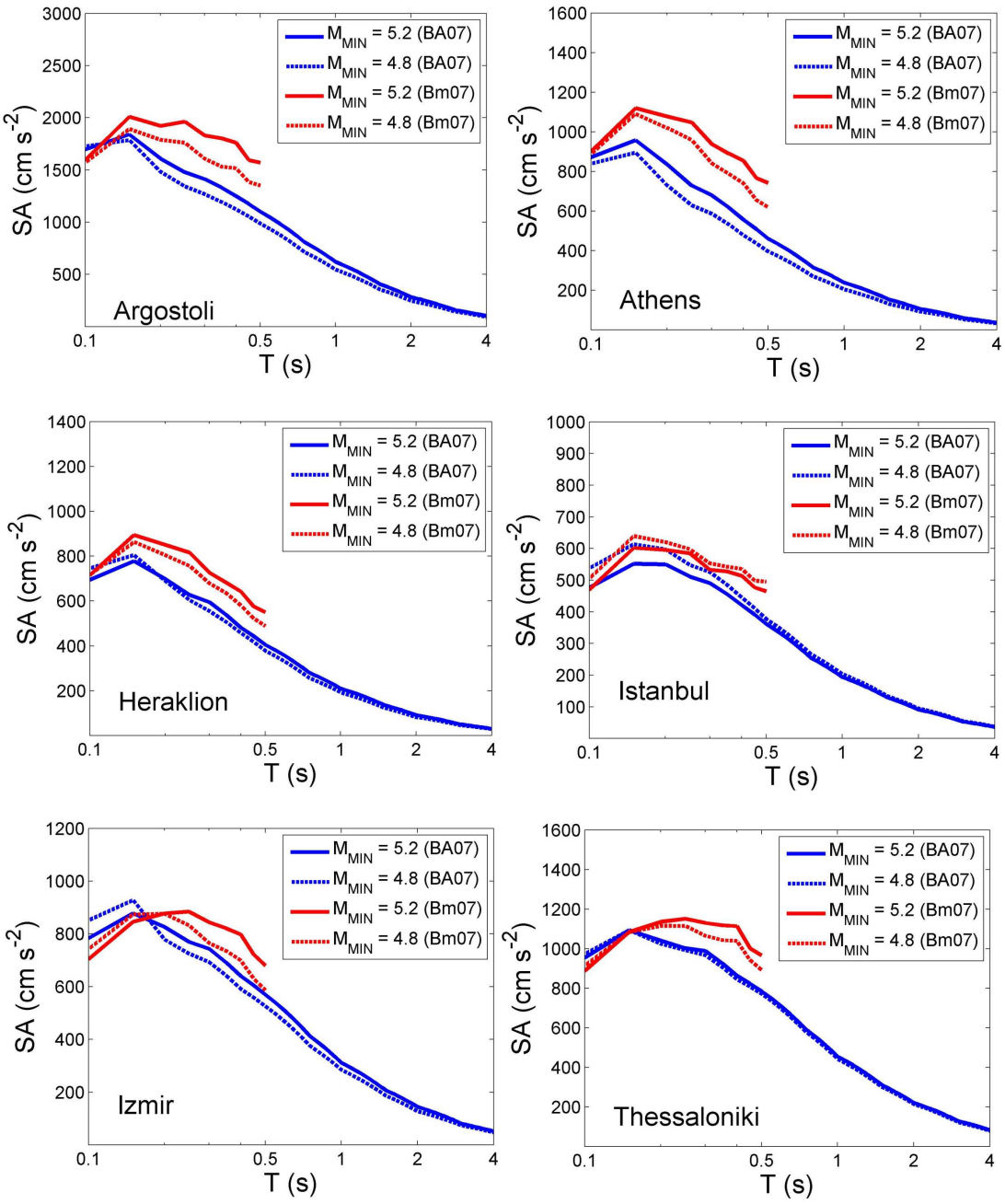

Figure 7.18: Comparison of UHS for M_{MIN} = M_W 5.2 (solid line), and M_{MIN} = M_W 4.8 (dashed line). Curves for the BA07 relation are shown in blue and Bm07 in red. PP2000 zone model and normal/strike-slip faulting assumed.

The comparisons of the two hazard analyses using different M_{MIN} are useful. They agree with observations of previous authors that indicate that a lower M_{MIN} increases the hazard at short periods. This is a good internal consistency check of the Monte Carlo method. It is also notable that in the case of Bender and Campbell (1989) and Grünthal and Wahlström (2001) the sensitivity analyses were conducted using the Cornell (1968)-McGuire (1976) method. This work, and that of Beauval and Scotti (2004), clearly demonstrates that the same trends are observed when the Monte Carlo method is used. Unlike the analyses of the aforementioned authors, this work encompasses areas of high seismicity and moderate seismicity. The comparison between areas of high seismicity and moderate seismicity suggests that the impact on M_{MIN} has a greater dependence on the overall seismicity level in a region than previous studies might suggest. This is worthy of investigation in future.

Although differences do exist in the hazard curves and hazard maps when different minimum magnitudes are compared the differences are small. They appear to be on the order of typically less than 30 $cm\ s^{-2}$, which for much of the Aegean is an order of magnitude less than the actual hazard value. This suggests that the minimum magnitude of M_W 5.2 chosen here is an appropriate choice for hazard analysis in the Aegean region.

7.3 Deep Earthquakes

7.3.1 Deep earthquakes affecting Seismic Hazard Analysis

The inclusion of deep earthquakes into the seismic hazard analysis raises several issues that have been touched upon in some of the previous sections. Following the example of previous Aegean hazard analyses (Papazachos, 1990; Papaioannou and Papazachos, 2000), the shallow source zones have been assumed to be homogenous to a depth of 60 km. For much of the Aegean this encompasses most of the observed seismicity, which can be attributed to the shallow seismogenic crust. The Hellenic arc is a subduction zone, however, and a non-trivial proportion of the seismicity of the South Aegean region is intermediate depth, which is attributable to this subduction.

The obvious question to ask is how do deep events contribute to hazard? The disaggregations shown in Chapter 6 indicated that for many areas of the Aegean, the hazard at a site is largely controlled by earthquakes within an epicentral distance of less than 40 to 60 kilometres (depending on the attenuation model used). This would suggest that although deep earthquakes may produce ground shaking at a site, their influence on PGA hazard may not be appreciable. Conversely, for seismic hazard analysis undertaken for the purposes of building large engineered structures, which will be affected by longer period acceleration, deep earthquakes may be significant.

The 2006 Kythera earthquake provided arguably the best illustration in recent times of

an intermediate depth earthquake resulting in damaging intensities (MMI VIII) being observed at the surface (Konstantinou *et al.*, 2006). Though the estimated hypocentre (60 km depth) would locate this event at the top of the intermediate-depth zone, it would suggest that a larger, deeper event may be equally capable of damage. This would lead to the conclusion that the impact of deep earthquakes needs to be explored, even if the eventual impact on the hazard analysis is trivial. In extending the hazard analysis beyond shallow seismicity there are several problems encountered. These need to be addressed in more detail.

7.3.2 Representation of deep earthquakes in the earthquake catalogues

One of the most poorly constrained parameters in the pre-ISC era period of the instrumental earthquake catalogue is hypocentral depth. The errors on the depth estimates may be on the order of tens of kilometres. Even for the early ISC period the likely depth estimates may still be erroneous by several kilometres. Depth estimates on historical earthquakes, which are derived from macroseismic intensity, are largely absent from the catalogue. The only exceptions to this are a small number of likely "deep" events where estimates to the nearest 25 to 50 km may have been attempted. Given these uncertainties, the common approach of selecting shallow earthquakes by introducing a hard cut-off for depths greater than a specified limit (a rather liberal 60 km here), risks excluding events that were actually within the shallow crust but misclassified as deep, and including events for which the opposite is true.

The input catalogue used in these analyses is considered complete above magnitude M_W 5.2 for the 1900 to 2005 period. This value is spatially stable in two dimensions, but not in three. Even for earthquakes used in estimation of the hazard parameters for many of the source zones, some have been arbitrarily assigned a depth of 15 km (the mean depth of shallow earthquakes in the Aegean) where depth estimates are missing. The ultimate consequence of this is that knowledge of earthquake depth has been limited even for earthquakes that have already been used in this analysis. For the purposes of estimating parameters of seismicity (rate, b-value etc) it is not obvious that there is any particular bias toward the exclusion of misclassified deep events or the inclusion of misclassified shallow events. Although there is considerable uncertainty here, there is little reason to assume that there is an intrinsic bias in the data set.

7.3.3 Deep source zones

A more significant problem encountered when incorporating deep events into the seismic hazard analysis, is the characterisation of the seismic source zone. In the existing source models, Papazachos (1990) recognised that two levels of deep seismicity could be delineated: shallow-intermediate depth ($60 \leq$ h (km) < 100) earthquakes, which are believed to represent the seismogenic layer at the coupling of the subduction African plate and

overriding crust, and deep-intermediate depth (h $\geq$ 100km) earthquakes, which are attributed to a more steeply dipping Benioff zone beneath the volcanic arc. This distinction is made more clearly in the intermediate-depth source zones developed by Papaioannou and Papazachos (2000). A similar distinction is made in the WT2006 source model developed in Chapter 3. There the shallow-intermediate depth zone and the deep-intermediate depth zone overlap when viewed from above.

Of the source models used, only the PP2000 model and WT2006 model have clearly defined and justified intermediate-depth source zones. The PZ1990 model does define source zones though albeit with little clear guidance as to how these zones have been developed. In all the models, there is little comment regarding the proportion of intermediate-depth seismicity found outside the defined zones around the Hellenic arc. Beyond the Benioff zone of the Hellenic subduction, deep seismicity is sporadic and well-distributed across the Aegean. Away from the cooler lithosphere of the subducting slab, the source mechanism for these deep events is less clear. It could be assumed that, like deep earthquakes observed in other regions, that these deep events may be attributed to heterogeneities in mantle convection, or solid-state mineral phase transitions (Frohlich, 1989), though these are usually associated with deeper focus events than those considered here. There is always the possibility that the depths of these anomalous events have been erroneously classified as deep events.

To incorporate intermediate-depth zones into the seismic hazard analysis several simplifications have been made. Generally there is good agreement between the source models that seismicity can be divided into three depth categories: shallow (depth $<$ 60 km), shallow-intermediate (60 $\leq$ h (km) $<$ 100) and deep-intermediate (depth $\geq$ 100 km). For the PP2000 and WT2006 models, uniform zones are assumed where delineated, and the remainder of deep focus seismicity is simulated by random re-sampling of the observed hypocentres and magnitudes at these depths. The uniform zones retain the assumption made in Papazachos (1990) that the slab interface is capable of very large magnitude events (up to M_W 8), whilst the maximum magnitude of events in the deeper Benioff zone is smaller. For the remaining source models, intermediate-depth seismicity is entirely simulated by random re-sampling (with replacement) of the observed seismicity. This maintains the spatial density distribution, but relaxes the assumption that only the shallow-intermediate depth events are capable of reaching very large magnitudes.

7.3.4 Strong Ground Motion attenuation from Intermediate-depth events

The greatest challenge in assessing hazard from intermediate-depth earthquakes is modelling the strong ground motion. Whilst an abundance of attenuation models exist for shallow earthquakes, there are very few that are widely used for intermediate-depth events. To this author's knowledge there are no ground acceleration attenuation relations in common use that have been derived from exclusively (or at least mostly) intermediate-

depth Aegean seismicity. The shallow crustal seismicity relations used to this point are not considered to be applicable beyond the 60 km depth limit assumed here. Many of them are derived using earthquakes no deeper than 30 km. This is in most part due to a deficiency of strong ground motion records from intermediate-depth earthquakes. Those that do exist are generally biased to the most active zones and dense strong motion networks (Japan, Taiwan, Alaska, Mexico etc.). Although the Hellenic arc is a subduction zone, it is not at all clear that attenuation relations derived for other subduction zones are applicable here. Nevertheless, the absence of Aegean based subduction zone attenuation models leaves no alternative but the import of external models.

An initial search of subduction zone attenuations was undertaken. Similar criteria as those set out in Cotton *et al.* (2006) were observed. Attenuation relations that are dependent on one region, such as Lin and Lee (2008), were rejected. Two attenuation models passed the initial selection criteria: Youngs *et al.* (1997) and Atkinson and Boore (2003) (erratum 2008). In both empirical relations the data set included earthquakes from Alaska, Chile, Cascadia, Japan, Mexico and Peru. For PGA attenuation the Atkinson and Boore (2003) attenuation relation was preferred over the Youngs *et al.* (1997) relation. This is for two reasons. The Atkinson and Boore (2003) relation (hereafter referred to as AB03) uses the same regression methodology as Youngs *et al.* (1997), but with a strong motion database containing four times as many records (each horizontal record is treated separately). It also considers non-linear magnitude scaling and site scaling. The second reason comes from a comparison of predicted ground acceleration and observed ground acceleration from the 2006 Kythera event using the Youngs *et al.* (1997) attenuation relation (Konstantinou *et al.*, 2006). This clearly showed that the Youngs *et al.* (1997) attenuation relation overestimated the PGA. Although no comparison is made for the Atkinson and Boore (2003) relation, the non-linear site scaling terms are likely to have improved this fit. However, there are not enough strong motion records from intermediate depth earthquakes from which a maximum likelihood fit (in the manner shown in Chapter 5) could be attempted.

The AB03 relation is derived using both horizontal records of ground motion separately. No conversion is made to scale this for equivalence to the geometric mean component. The functional form of this relation for PGA is:

$$\log Y = c_1 + c_2 M + c_3 h + c_4 \left(\sqrt{R_{RUP}^2 + \Delta^2}\right) - \gamma \log\left(\sqrt{R_{RUP}^2 + \Delta^2}\right) + \cdots$$
$$\cdots + c_5 slS_C + c_6 slS_D + c_7 slS_E \pm \sigma \tag{7.1}$$

where Δ is a near source saturation term given by $\Delta = 0.00724 \times 10^{0.507M}$, γ is a scaling term defind as $\gamma = 10^{(1.2-0.18M)}$ for interface events, and $\gamma = 10^{(0.301-0.01M)}$ for inter-slab

events. The sl term represents a non linear site scaling defined (for PGA) as:

$$sl = \begin{cases} 1 & \text{if} \quad PGA_{rx} \leq 100cm\ s^{-2}, \\ 1 - (PGA_{rx} - 100)/400 & \text{if} \quad 100 < PGA_{rx}\left(cm\ s^{-2}\right) < 500, \\ 0 & \text{if} \quad PGA_{rx} \geq 500cm\ s^{-2}. \end{cases} \tag{7.2}$$

where PGA_{rx} is the predicted PGA on rock (NEHRP B). All the coefficients are dependent on whether an interface or intra-slab event is assumed; hence two different formulae are used. The two attenuation functions (interface and intra-slab) are plotted in Figure 7.19.

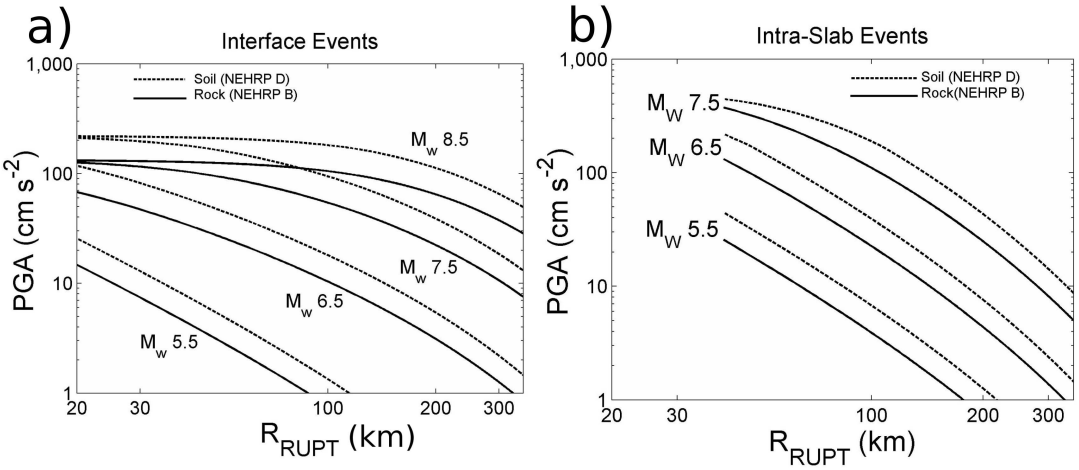

Figure 7.19: The AB03 PGA attenuation relation for interface (a) and intra-slab (b) events on a rock site (solid line) and soft soil site (dashed line).

The AB03 relation does also contain a small number of spectral parameters. As these relations are derived from intermediate depth earthquakes the spectra are more intensely sampled at longer period accelerations than those relations for shallow earthquakes. Furthermore, an erratum to this paper was issued (Atkinson and Boore, 2008) that highlights an error in the regressions for the 2.5Hz and 5 Hz attenuations. Only the PGA form of the relation will be used here.

7.3.5 Deep earthquakes in the Monte Carlo Seismic Hazard Analysis

The seismic hazard analysis incorporating intermediate depth earthquakes is shown in Figure 7.20. The MCMAEU method is used here on a rock site. Several simplifications have been made in the implementation of the AB03 attenuation relation. The AB03 model uses rupture distance as the preferred metric. Given the depth of the earthquakes considered (h $>$ 60 km) this is assumed equivalent to hypocentral distance. This is simply because for all but the largest earthquakes, of which there are very few, the hypocentral distance will be substantially greater than the rupture length. Although the distinction between interface and intra-slab events is clear in the attenuation model, it is less so in the source model. A crude simplification is made where the shallow-intermediate depth

events ($60 \leq$ h (km) < 100) are assumed to be interface events, whilst deep-intermediate depth events (Depth ≥ 100 km) are assumed to be intra-slab events. The former is a reasonable approximation for the Hellenic arc. The latter assumption is not entirely accurate as events in this range may be interface events. There will be an increasing proportion of intra-slab events at these depths, so in lieu of more specific information this approximation is reasonable.

Figure 7.20 clearly shows that when PGA is the hazard variable being considered, the inclusion of deep earthquakes appears to have little impact on the overall hazard. This is borne out by the percent change in hazard shown in Figure 7.21. The impact is clearly variable from site to site, with most changes in hazard being no more than 5 %. In some locations a slight overall decrease in hazard is observed. A proportion of this may be due to the natural variability of the Monte Carlo process, and some due to the attenuation model and its aleatory variability term. It should also be noted that the areas showing the greatest increase or decrease in hazard are mostly in lower hazard regions and that the differences in hazard are typically on the order of no more than 10 to 20 $cm\ s^{-2}$.

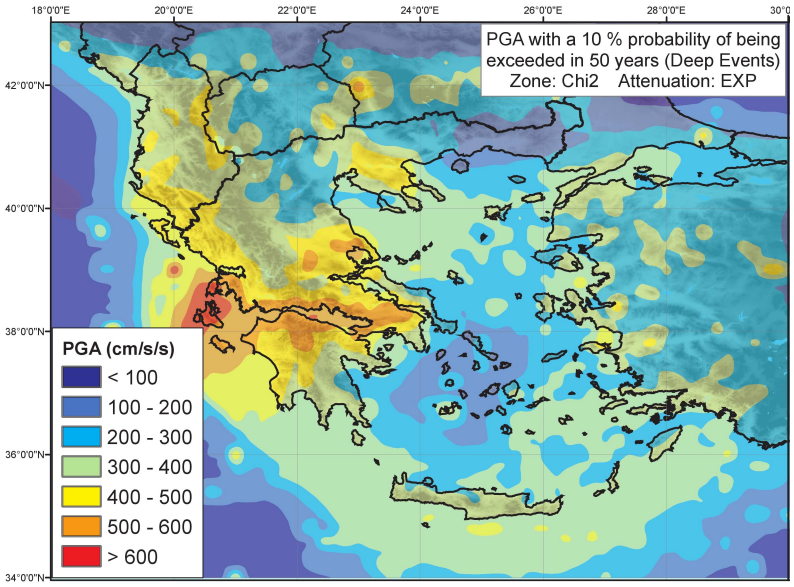

Figure 7.20: PGA with a 10 % probability of being exceeded in 50 years with inclusion of deep events (χ^2 zone model weighting, EXP attenuation model weighting, rock site, normal/strike-slip faulting)

Given the inherent problems of estimating hazard from deep earthquakes it is not surprising that the overall impact on the hazard analysis is not that great. As with many of the modifications to the hazard analysis covered in this chapter there is still a considerable amount of uncertainty that cannot be incorporated into the model. The absence of an Aegean-specific (or even Aegean inclusive) subduction attenuation relation means that a relation has to be imported from elsewhere. Whilst the number of strong motion records from intermediate depth earthquakes in the Aegean is increasing, the data set is still too small to derive robust regression equations. Nor can they be used to assess the fit of imported attenuation models to observed data quantitatively. In addition, the im-

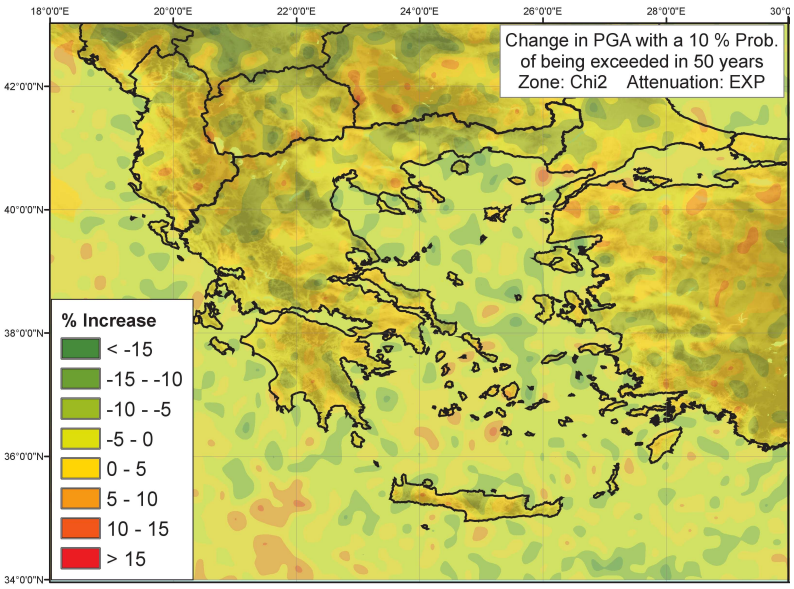

Figure 7.21: Percentage change in PGA with a 10 % probability of being exceeded in 50 years when deep earthquakes are included in the hazard analysis.

pact that deep earthquakes may have on a hazard analysis should be greatest at longer spectral periods of acceleration. As with many strong motion databases, the number of records from which regressions of spectral acceleration can be achieved at long periods is smaller still. It can only be hoped that as the coverage of the strong motion networks in the Aegean expands, thus providing better strong motion records from intermediate depth events, the ground motion attenuation models can be better constrained.

7.4 Incorporating Site Condition in a Seismic Hazard Map

7.4.1 Separate site condition maps

All of the maps shown in this work so far have assumed a "rock" site. This is defined as being NEHRP Class B, with an assumed V_{s30} of 760 $m\ s^{-1}$. The assumption of a rock site is a common one for mapping purposes as it represents a "shallow bedrock" material (or close approximation thereof). One of the most important differences between a hazard map and a site-specific seismic hazard analysis is in the characterisation of site. It is now standard practice to see strong ground motion attenuation relations incorporate a site characterisation term, be it in the form of a simple site parameter (e.g. 0 for rock, 1 for stiff soil, 2 for soft soil etc) or some characteristic property of the site material (e.g. V_{s30}). All of the attenuation relations used in these hazard maps have a "site" term. For site-specific analyses the soil type is easily input into the attenuation relation. For mapping purposes this is more difficult as, until recently, site conditions cannot be resolved over the sort of spatial scale that is needed for seismic hazard mapping.

Differences in site condition are commonly included in a seismic hazard analysis by plot-

ting different hazard maps for different soil types. Typically this breaks down into three categories: rock, stiff soil and soft soil. These categories generally reflect the extent to which site condition can be resolved in the strong motion data set used in the construction of the attenuation relations. An illustration of this can be seen in Figure 7.22, which shows the differences in seismic hazard for the 10 % probability of being exceeded in 50 years. Here, maps are plotted for the DT07 and Bm07 relation with an assumed rock, stiff soil and soft soil type.

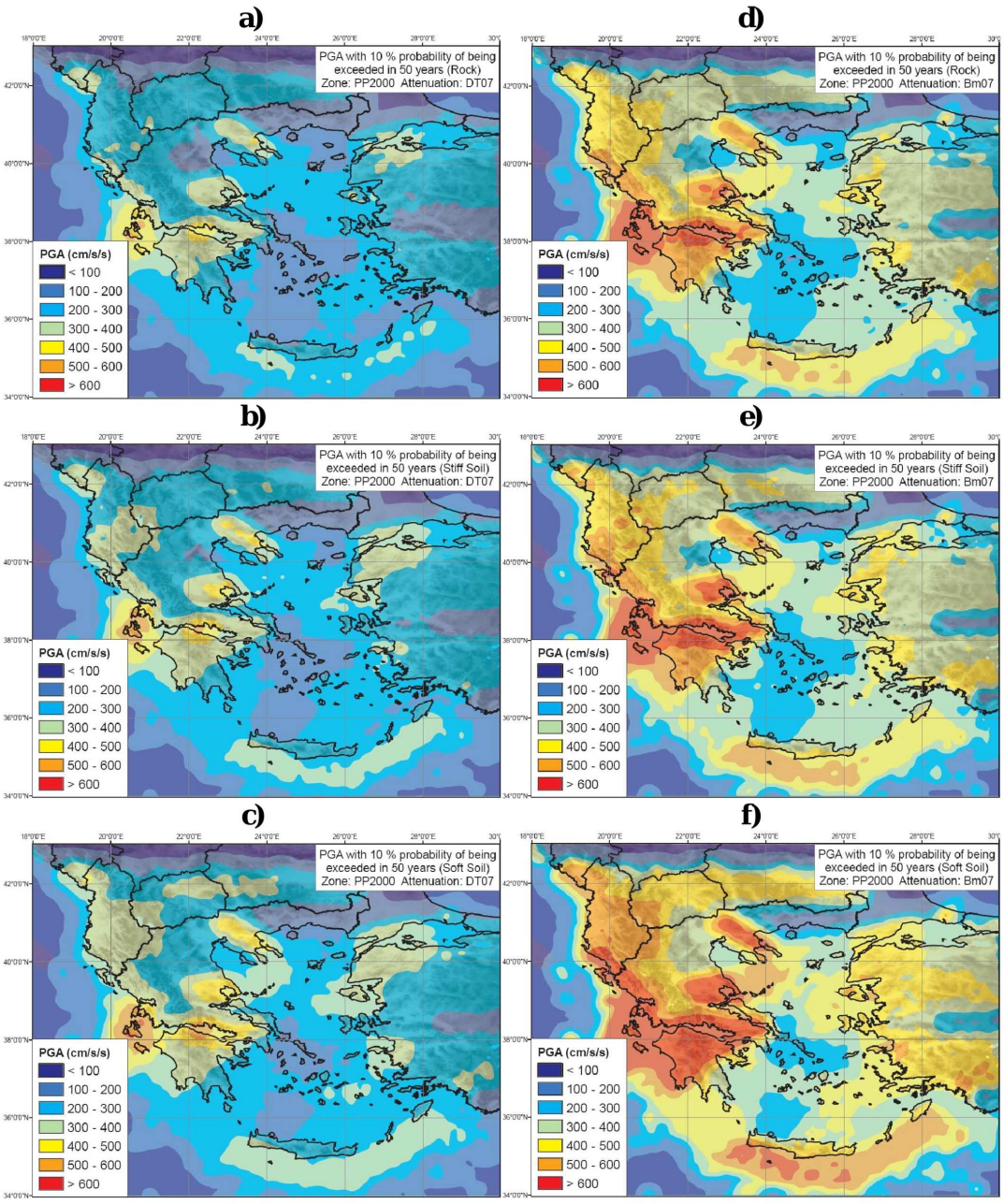

Figure 7.22: PGA with a 10 % probability of being exceeded in 50 years (PP2000 source model, normal/strike-slip faulting) on a rock site (a and b), stiff soil site (c and d) and soft soil site (e and f). DT07 attenuation model used in a), c) and e), Bm07 in b), d) and f).

These maps are not especially useful as they don't demonstrate the regional variation in

hazard due to site condition. Instead the combinations of maps only serve to illustrate the fact that the attenuation relations predict higher PGA on stiff soil and soft soil sites than they do on rock sites. What is needed therefore is a means of characterising spatial variation in site condition that can be practically integrated into seismic hazard mapping calculations.

7.4.2 A Global V_{s30} data set

The estimation of V_{S30} for a particular site is a complicated process, often requiring detailed geotechnical investigation to develop a profile of shear wave velocity variation with depth beneath the site. To undertake a spatial investigation of V_{S30} by direct measurement, to such a resolution that it could be practically used for seismic hazard mapping applications, would be an impractical and expensive process. To map site condition effectively, a proxy measure relating it to a spatially contiguous variable is needed. The most obvious source of information is surface geology. If characteristic shear wave velocity profiles exist for a large number of sites, which effectively sample the regional variation in surface geology, then a map of V_{S30} can be assembled according to the distribution of different geological units. Such a map for California was produced by Wills and Clahan (2006). Here, 19 different geological units were classified, based on 556 shear wave velocity profiles and geological observation. For each unit an average profile, based on the composite of all profiles classified as belonging to that unit, is constructed. From this average profile the mean V_{S30} and its standard deviation are calculated. Using geological maps and assigning mean V_{S30} to each geological unit a map of V_{S30} was assembled.

Mapping V_{S30} by geological proxy is a theoretically robust method of analysing distribution of site condition. However, there are many sources of error. Shear wave velocity, even within a geological unit, will be influenced by factors such as the depth of sediment, the porosity, grain site and degree of consolidation. Heterogeneities in all these quantities can vary shear wave velocity within the geological unit; hence the necessity to define a mean and standard deviation of V_{S30}. Given this uncertainty, rather than produce a map of V_{S30}, a map of NEHRP site condition is shown in Wills and Clahan (2006). Where uncertainty in V_{S30} could result in different site classes being considered, a borderline category NEHRP Class B/C or C/D is defined.

Surface geology as a V_{S30} proxy is useful, but it depends on having good knowledge of the surface geology of a region and requires a large number of profiles. Where money and resources are available to do this then geologically based V_{S30} may be possible. For much of the globe, however, this is not possible. Instead a different, but closely correlated site condition proxy is used.

Wald and Allen (2007) have produced the first global database of V_{S30} using topographic slope as a site condition proxy. Using a total of 1197 V_{S30} measurements taken across three

active continental margins (California, Italy and Taiwan), and a further 520 from two stable continental interiors (Memphis and Australia), empirical relations between V_{S30} and topographic slope (determined from the SRTM 30-sec global database) have been determined. Allen and Wald (2007) indicate that the values for V_{S30} given in the online database (http://earthquake.usgs.gov/research/hazmaps/interactive/vs30/ - accessed April 2008) are not derived directly from regression. Instead the V_{S30} values are binned into narrower windows according to NEHRP site class. V_{S30} is then calculated by interpolating between the bin edges according the degree of topographic slope. Separate regressions are undertaken for active tectonic regions and stable continental interiors. Here the Aegean is classified as an active tectonic margin.

The decision to disseminate this database worldwide has meant that hazard mapping using V_{S30} can now be widely applied. Before this is done, it is important to assess whether topographic slope is an appropriate proxy for V_{S30}. The rationale behind this method lies in the geotechnical conditions of soil and rock and their impact on shear-wave velocity. In soils, changes in shear modulus are mostly attributed to void ratio. Since density and mean effective principal stress do not vary greatly in the upper 30 m of soil, these variables do not have a great influence at shallow depths. Void ratio is heavily controlled by grain size, and it is found that shear wave velocity increases with increasing grain size. In depositional environments (characterised by low topography), particle size decreases due to the lower energy available from fluvial mobilisation. In rock conditions a well-consolidated hard rock with coarse fracture spacing will have higher shear wave velocities. These same rocks will also be more resistant to weathering allowing them to hold a steeper slope.

This underlying theory helps explain why V_{S30} correlates with topographic slope. There are, however, many environments where these assumptions break down. These include glacial terrain and volcanic plateaus. Comparison of the V_{S30} data for California with the Wills and Clahan (2006) map shows good correspondence between the two methods. Topographically derived V_{S30} appears to be more spatially variable than those derived from geology. However, since geologically derived V_{S30} maps don't take into account variation in material properties, such as grain size, within a single geological unit, an argument can be made that the topographically derived V_{S30} values are more adept at capturing variation in V_{S30}. The added advantage is that V_{S30} can be mapped uniformly across the globe and does not require geotechnical investigation of particular regions. The global V_{S30} database is now used in the USGS Shakemap program.

The online global V_{S30} server was accessed to obtain V_{S30} maps for the Aegean region. The original V_{S30} data is shown in Figure 7.23a, and NEHRP site class in Figure 7.23b. The latter is derived using the V_{S30} boundaries indicated in Wald and Allen (2007). In the implementation of hazard calculations no site amplification is assumed for offshore locations.

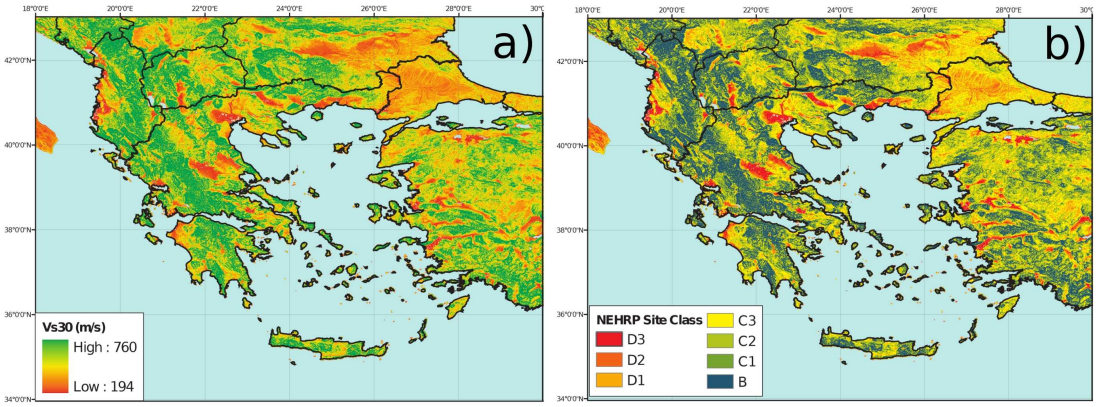

Figure 7.23: V_{S30} maps for the Aegean region. a) Continuous scale ($m\ s^{-1}$), b) NEHRP categories.

7.4.3 How to use the Global V_{S30} data set in hazard analysis

With a global data set of V_{S30} estimates available, how should they be integrated into the hazard analysis?. The most obvious means of integration would be the calculation of hazard for each grid point in the V_{S30} data base. Over a small area such as a city or a province this is a feasible approach. It could be argued though that variation in seismic hazard cannot be truly resolved on such a fine scale. This begins to blur the line between mapping and site-specific analysis. Whilst site condition can change enormously over distances of no more than one or two kilometres, is it likely that the underlying seismic hazard will vary to that degree? With errors on observed epicentral location of the order of 1 to 2 km, and the uncertainty in fault location and conversion between distance metrics, it seems unlikely that hazard for two sites spaced only one kilometre apart will have a different seismic hazard, before site condition is taken into account. This illustrates a concept that could be colourfully referred to as a "weather map fallacy". If the processes underlying the calculation of the variable being mapped cannot be adequately modelled or resolved at the resolution of the map, then the presentation of a high resolution map serves little purpose beyond aesthetic appearance.

The other argument against calculating hazard at the same resolution as the V_{S30} data set is computational practicality. The number of data points for hazard calculation is the greatest influence on computational time. Synthetic catalogues of considerable length are generated in a manner of seconds, but the calculation of hazard on a $0.2° \times 0.2°$ grid takes nearly 50 times as long as calculation on a $0.5° \times 0.5°$ grid. Calculation of hazard at 30 arc-second resolution would take an inordinate amount of time. Clearly interpolation is needed.

All the maps shown previously have been interpolated onto a finer grid, assuming a homogenous site type. For the purpose of contouring, this is inevitable and some discussion on this has been given in chapter 6.2. As it is necessary to interpolate the hazard calcula-

tions onto a finer grid for displaying the map, there is little additional cost to interpolating the hazard calculations onto the V_{S30} grid.

The problems arise when considering how to then incorporate the V_{S30} data into the hazard calculation. This is where the characterisation of site condition in the selected attenuation relation plays an important role. In the DT07, Bm07 and Am05 relations, $\log Y$ scales linearly with the NEHRP site classification term. This means that site condition can be addressed for each site on the V_{S30} grid by simply taking the logarithm of PGA hazard at a site, adding the site parameter depending on site classification, then taking the exponent to return to PGA. For Am05 this becomes:

$$PGA_{SITE} = 100 \times 10^{(\log_{10}(PGA/100)+0.05SA+0.137SS)} \tag{7.3}$$

where SA = 1 for stiff soil (0 otherwise) and SS = 1 for soft soil (0 otherwise). For DT07:

$$PGA_{SITE} = 10^{(\log_{10}(PGA)+0.038S)} \tag{7.4}$$

where S = 0, 1 or 2 for rock, stiff soil and soft soil respectively. For Bm07:

$$PGA_{SITE} = 10^{(\log_{10}(PGA)+0.0208SA+0.0781SS)} \tag{7.5}$$

where SA = 1 for stiff soil (0 otherwise) and SS = 1 for soft soil (0 otherwise). In these relations site is classified as rock, stiff soil or soft soil according to V_{S30}, using the NEHRP classifications given previously.

For attenuation relations with non-linear scaling of PGA with site there is additional complication. This is true of the BA07 relation for which $PGA_{SITE} = f\,(PGA, pga4nl, V_{S30})$, where pga4nl is the PGA determined for a reference rock site ($V_{S30} = 760\ m\ s^{-1}$) with $\sigma = 0$. To scale PGA according to site condition, both the original PGA data and the pga4nl data for the hazard calculation must be input. Otherwise, the procedure is similar to that for the linear site scaling relations, albeit with a non-linear scaling term.

The maps shown in Figure 7.24 indicate the impact that inclusion of V_{S30} has upon a single model hazard map. Here the PP2000 zone model and BA07 attenuation model are used. A smoothed hazard map (Figure 7.24a) is compared with a contoured hazard map (Figure 7.24b) to elucidate the effects of the site condition. The smoothed hazard map clearly demonstrates how geographical features such as river valleys and low lying alluvial plains have a raised seismic hazard. In the contour map this has the effect of slightly blurring the boundaries between different hazard levels in accordance with the topography. The areas where seismic hazard is raised most significantly are in the Thessaly province of eastern Greece and the valleys of the Cayster and Büyük Menderes rivers in western Turkey.

The incompatibilities between the linear site scaling attenuation relations (Am05, DT07

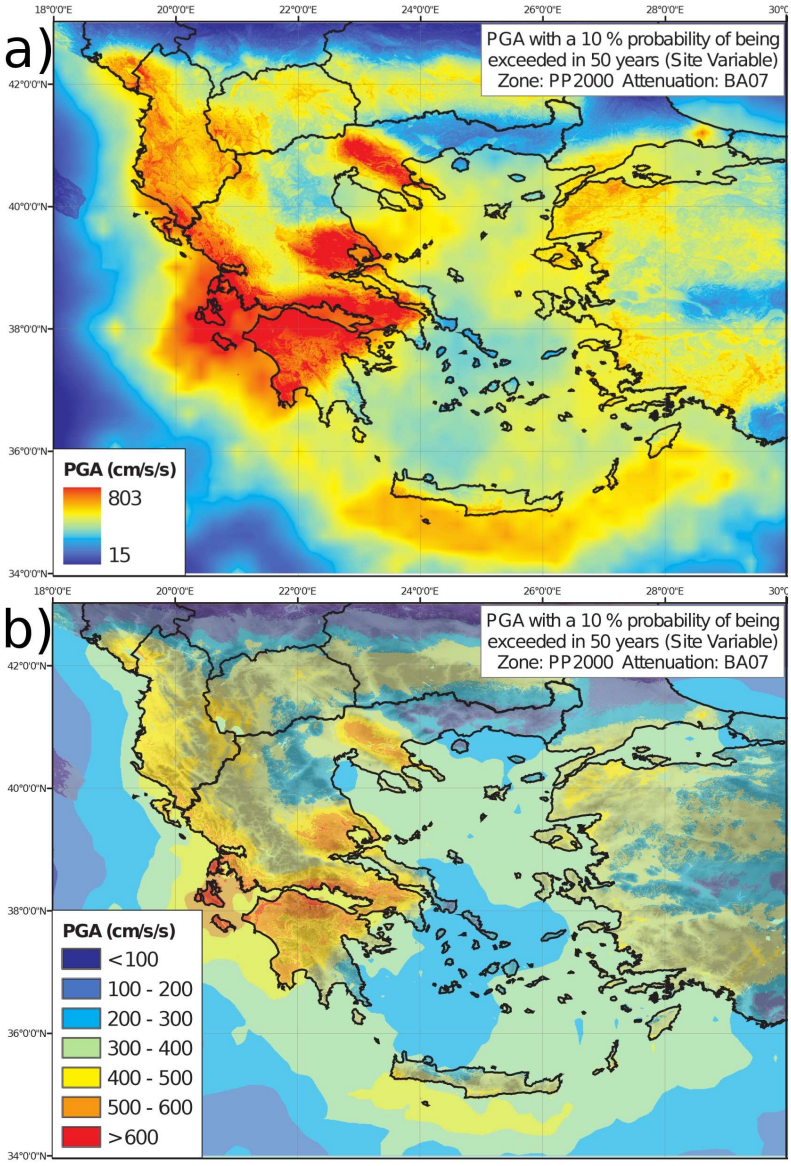

Figure 7.24: PGA with a 10 % probability of being exceeded in 50 years with variable site condition. PP2000 zone model and BA07 source model used and normal/strike-slip faulting assumed. a) Continuous scale ($cm\ s^{-2}$) and b) Contoured.

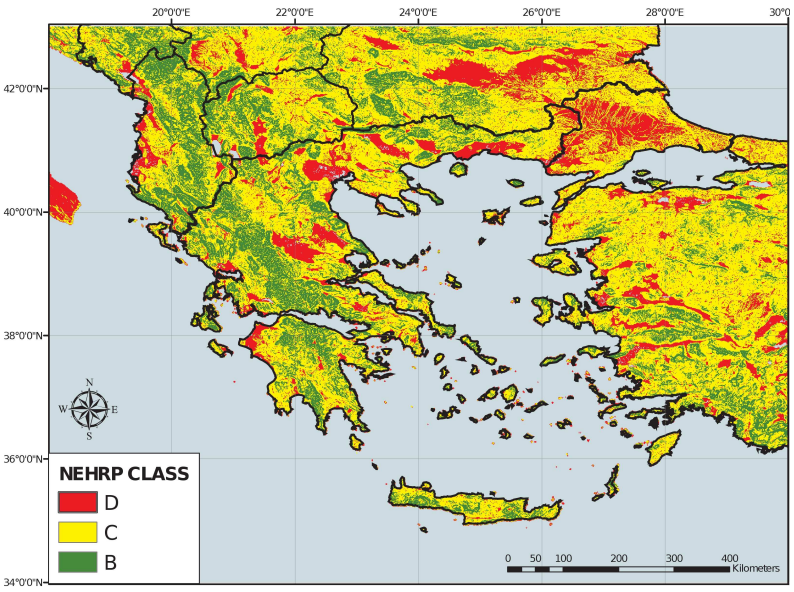

Figure 7.25: Site categorization across the Aegean region using 3 NEHRP classes.

and Bm07) and BA07 make integration of site condition in an epistemic uncertainty analysis much harder. This is complicated by the fact that Am05, DT07 and Bm07 all characterise site according to NEHRP site class, whereas BA07 uses V_{S30} directly. It is therefore not possible to incorporate V_{S30} directly into the MCMAEU unless some approximations are made. Since Am05, DT07 and Bm07 all consider site effect in terms of three NEHRP classes (B, C and D) the V_{S30} data set is divided into the same three classes, shown in Figure 7.25. This allows Am05, Bm07 and DT07 to be compared. To incorporate BA07, the V_{S30} data cannot be used directly; instead a characteristic V_{S30} value is assigned to each site class. Following suggestions in Stafford *et al.* (2008b), an "intermediate" V_{S30} was assigned to each site based on its site class (NEHRP B = 760 $m\ s^{-1}$, NEHRP C = 570 $m\ s^{-1}$ and NEHRP D = 300 $m\ s^{-1}$). This reclassification is rather crude but it does allow for reasonable comparison between NEHRP site classes and V_{S30}. Given the uncertainties inherent in the estimation of V_{S30} via topographic proxy, this reclassification may actually be more appropriate in this environment.

The maps produced by combining the MCMAEU method with a variable site database (Figure 7.26), based on NEHRP Class, rather than V_{S30}, display many of the same characteristics as those maps derived from a "single model" method. These include amplification along river valleys and low lying plains in eastern Greece. This would suggest that the V_{S30} data can, by virtue of a reclassification into NEHRP site class, be incorporated into the MCMAEU method and its impact directly assessed.

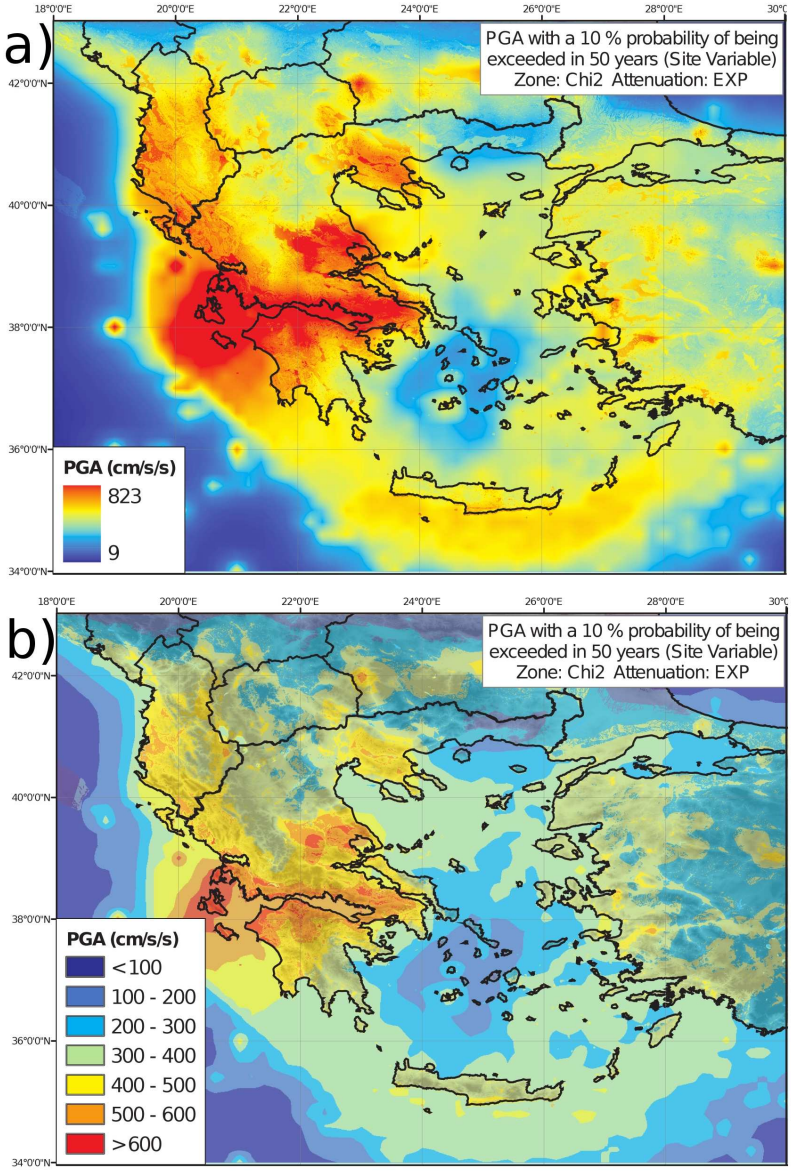

Figure 7.26: As Figure 7.24, using the MCMAEU method with χ^2 zone model weightings and EXP attenuation model weightings. a) Continuous scale ($cm\ s^{-2}$), b) Contoured.

7.5 Incorporating Fault Type into the Hazard Maps

7.5.1 The need to consider fault type

A key development of strong ground motion attenuation models in the last 10 years has been the inclusion of a "style-of-faulting" term. This has been introduced following recognition that the type of fault rupture has an influence on the strength of ground motion arising from the rupture. Thrust and reverse faults tend to produce more severe ground shaking than strike-slip events (Boore and Fumal, 1997; Campbell, 1997). Whilst the distinctions are less obvious, compounded by the lack of strong motion records from normal faulting events in California, it would also appear that normal faulting events are generally associated with slightly weaker strong ground motions than strike-slip or thrust events (Skarlatoudis *et al.*, 2003; Danciu and Tselentis, 2007). Spudich *et al.* (1999) refine this generality and observe that strong motion from both strike-slip and normal events in extensional regimes is lower than that in compressional regimes. These observations have been largely borne out in the Aegean region (Margaris and Hatzidimitriou, 2002; Skarlatoudis *et al.*, 2003; Danciu and Tselentis, 2007).

As it is recognised that fault type does have an influence on strong ground motion, this has been incorporated into attenuation models. The seismic hazard maps shown previously have assumed normal or normal/strike-slip faulting, depending on how they have been defined in the attenuation relation. The diverse seismotectonics of the Aegean region has already been discussed in earlier chapters. Whilst normal/strike-slip faulting may be a reasonable assumption for much of Greece and western Turkey away from the Hellenic Arc and Adriatic coast, it cannot be assumed across the entire Aegean region. In the absence of well-constrained fault locations, variation in faulting is difficult to integrate into the hazard analysis, especially with the use of the Monte Carlo method.

There is little successful precedent for the inclusion of fault type into seismic hazard analysis via Monte Carlo simulation. One of the more sophisticated approaches is implemented by Sinadinovski *et al.* (2005) for southwest Western Australia. Here, synthetic ruptures are simulated with each rupture given a magnitude (in the same manner used here), a length (derived from magnitude using the empirical relations of Wells and Coppersmith (1994)), an azimuth, a dip and a dip angle. Initially the rupture azimuths are simulated from a uniform distribution between 0° and 360°. They are then rotated into alignment with the fault zone in which they are simulated. Dip angle is defined by assigning half of the dips to the left of the rupture, the other half to the right of the rupture. It should be recognised that in terms of seismicity, southwest Western Australia is not comparable to the Aegean. Being a region of low-moderate seismicity, only four zones are defined (plus background), all of which display mostly reverse faulting with E-W pressure axes. This is in stark contrast to the Aegean where fault type and alignment of P- and T-axes changes across the region. In many respects the approach by Sinadinovski

et al. (2005) is attempting to turn areas of epistemic uncertainty into aleatory variability by using the Monte Carlo method to define rupture planes. The relative seismotectonic homogeneity of Western Australia may mean that this assumption is reasonable. In the Aegean it is not.

7.5.2 Characterisation by simple fault type

Arguably the simplest approach is to assign each zone within a source model a "characteristic fault type". For the PP2000, PZ1990 and WT2006 models, each zone is assigned as normal, strike-slip, thrust or oblique. No distinction is made between oblique faulting in an extensional context and oblique faulting with a thrust component. This assignment is done manually and requires a reasonable knowledge of Aegean seismotectonics. As a crude representation of the seismotectonic variability of the region, such an approach may not be unwarranted. The degree to which faults need to be defined is limited by the parameters required for input into the attenuation relations. The five relations used here require only fault type as an input parameter. Other relations, including some of the other NGA models, may require further parameters such as depth to the top of the rupture.

There are many reasons to avoid defining random ruptures in the manner of Sinadinovski *et al.* (2005). Firstly, the ambiguity in dip angle cannot be fairly resolved by an even split between left and right of the surface rupture trace. Secondly, the required distance metrics are a mixture of epicentral distance and Joyner-Boore distance. It is necessary to define the location of the hypocentre of the rupture plane, which is in itself a further cause of uncertainty. In these circumstances it is preferable to use an empirical relation to derive Joyner-Boore distance from epicentral distance, as the known scatter in the empirical relation can be incorporated into the aleatory variability of the attenuation relation (Bommer *et al.*, 2005). Ultimately there are too many uncertainties in the features of the rupture planes of observed earthquakes in the Aegean to provide a basis for simulating synthetic rupture planes.

For the PP2000, PZ1990 and WT2006 models a characteristic strike, dip and rake are defined for each zone, including uncertainties ($1\sigma = 10^\circ$ for strike and rake; 3° for dip). These values have been determined from the literature pertaining to the models themselves, with further judgements assisted by the Aegean rupture catalogue previously used in Chapter 4 (Papazachos *et al.*, 1999) and the Atlas of Isoseismal Maps (Papazachos *et al.*, 1997). Normal faulting with an approximately E-W strike is assumed for all seismicity outside of the uniform zones. The fault classification for the zones in the PZ1990, PP2000 and WT2006 models are shown in Figure 7.27.

This crude assignment of faulting parameters begins to resolve the influence of regional scale variations of fault type in the seismic hazard analysis. It allows for distinction

between areas of compressional, extensional and transform faulting, which given the equally crude representation in the attenuation models is adequate for this purposes. It does, of course, mask the much more localised variation in fault type, including branch faults and faults representing transitions between seismotectonic regimes. It also fails to make any distinction between differences in faulting mechanism owing to asperities in the rock. For a site specific hazard analysis these small-scale variations in fault type may be influential. There is, however, a limit as to the detail that is resolved in the input model for seismic hazard analysis.

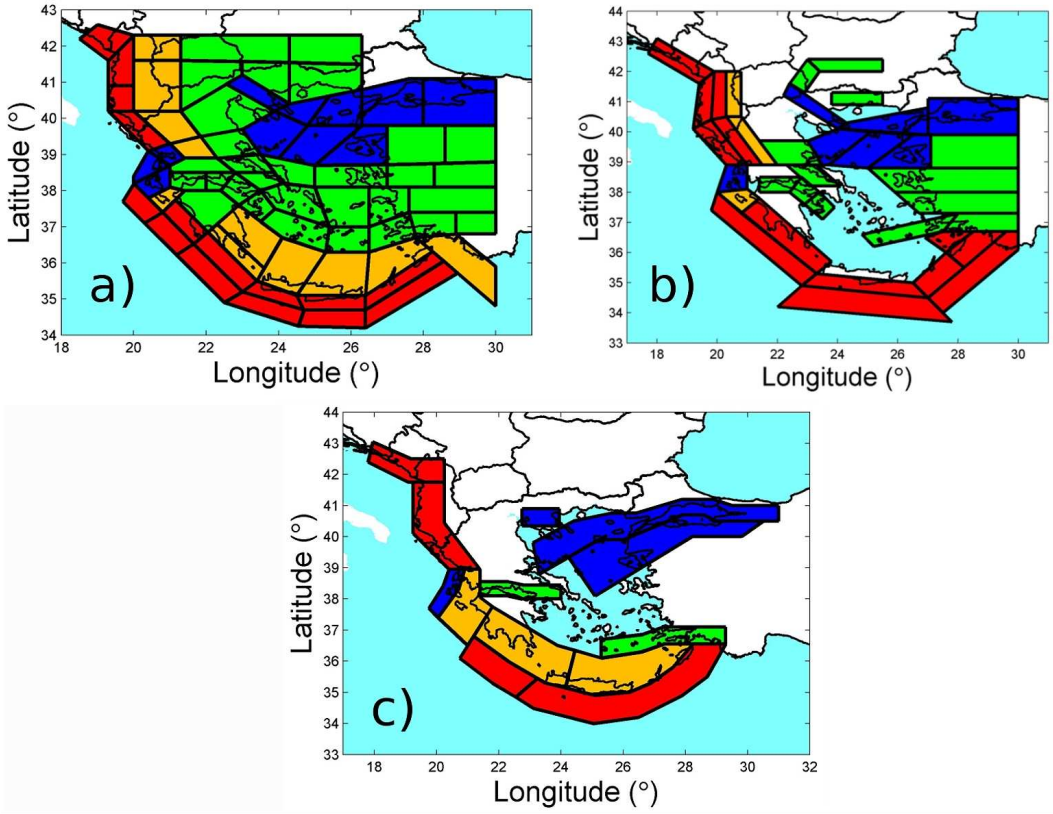

Figure 7.27: Assigned fault types for the a) PP2000, b) PZ1990 and c) WT2006 model. Thrust faults shown in red, strike-slip in blue, normal in green and mixed faulting in orange. The background is assumed to be normal faulting.

7.5.3 Fault characterisation in the K-means derived zones

A particular convenience of the zones derived by K-means partition of the Aegean rupture catalogue (Chapter 4) is that each zone will already contain a set of observed rupture parameters within it. Rather than assigning a characteristic rupture to each zone; strike, dip and rake can all be randomly sampled (with replacement) from the subset of ruptures in each zone. As before, some Gaussian scatter is added to each parameter to allow for uncertainty in the focal mechanism. In theory, this same approach could be used by partitioning the rupture catalogue according to the PP2000, PZ1990 and WT2006 models. The problem in doing so is that some of these zones may not contain any ruptures from

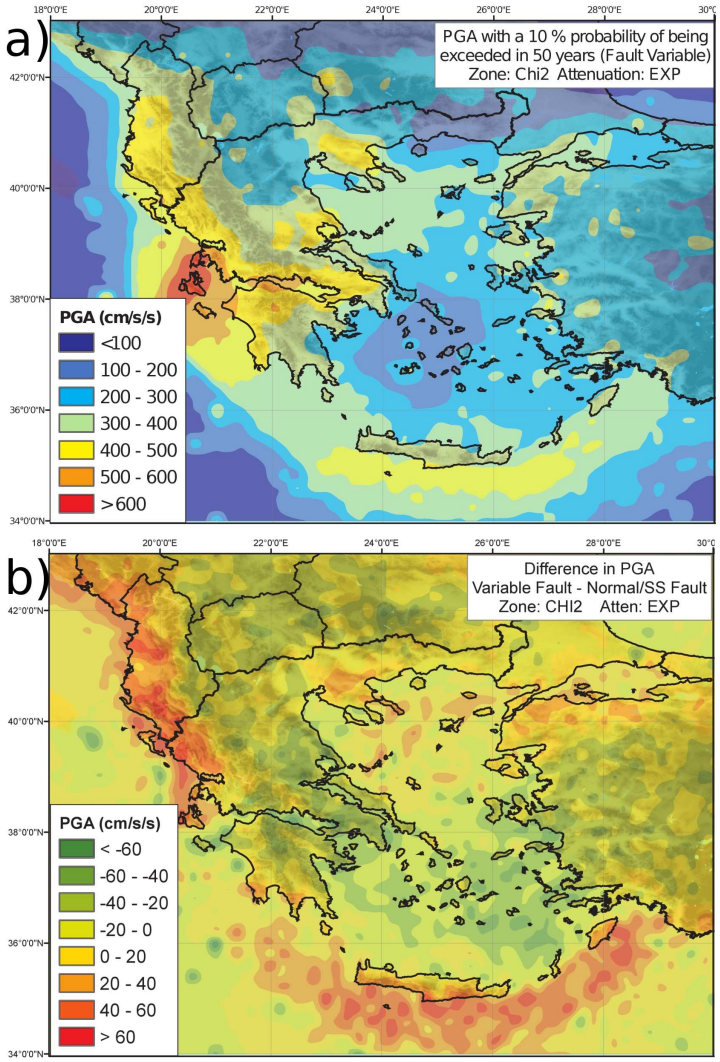

Figure 7.28: a) PGA with a 10 % probability of being exceeded in 50 years using the fault variable model, b) Change in PGA hazard from the assumed normal/strike-slip fault

the catalogue, thus requiring assignment of a characteristic rupture anyway.

The particular advantage of this method is that the real variation is faulting is captured more adequately in the Monte Carlo models. Some zones, particularly in regions of transition from thrust to strike slip faulting (south Ionian Islands) or strike-slip faulting to extension (North Aegean Sea), display different fault types over a smaller area. In the K-means derived zones, both types of faulting can be simulated in the zone, in proportion to their observed occurrence. The assumption of zone uniformity means that both types of fault can be simulated anywhere in the zone. In the most active regions where focal mechanism can vary over small spatial scales (such as the Ionian Islands) this is a reasonable supposition.

Figure 7.28a shows a seismic hazard map using the MCMAEU method, allowing for fault type to vary in accordance with the models. The differences in PGA between the variable fault and the assumed normal/strike-slip fault map (previously seen in Figure

7.26a) are shown in Figure 7.28b. This clearly demonstrates that although the scale of difference is relatively small (less than 20 $cm\ s^{-2}$ for more than three quarters of the sites considered) the incorporation of variable faulting does have an appreciable difference. As expected, those regions associated with thrust faulting along the Eurasian plate margin display higher hazard when faulting term is introduced. This is true, to a lesser extent, in northwest Turkey and the North Aegean Sea. Conversely, hazard is lower for much of mainland Greece and southwestern Turkey, where tectonic extension is dominant.

A reasonable question to ask is whether it is really necessary, for the purposes of hazard mapping, to attempt to constrain all the parameters required by the attenuation relations. The spatial variation in site condition and fault type is only crudely characterised by the approximations made here. True variation can occur on scales considerably smaller than the resolution of these models, which makes the true hazard considerably more variable than shown in these maps. This illustrates the differences between site-specific analysis and hazard mapping. It is crucial within a site specific hazard analysis that the site condition is well-determined and that the nature of active faults contributing to hazard at the site well-understood. The Monte Carlo method of seismic hazard analysis can accommodate this information easily; hence its increasing use in such applications. Again, it needs to be re-iterated that the process of seismic hazard mapping requires different inputs from site-specific analysis and cannot be considered as simply site-specific analysis repeated over a large grid of sites. What the last two sections of this chapter have shown is that even crude models of fault and site condition can have an appreciable impact on the seismic hazard map. They represent convergence toward an "idealised" map that can be produced to show the true variation in hazard, rather than a scenario map that considers only a single rock type or fault type.

7.6 Including Non-Poissonian Events in Seismic Hazard Analysis

7.6.1 Incorporating Non-Poissonian Events

The seismic hazard analyses shown so far are examples of time-independent seismic hazard analysis. This assumes that the rate at which a level of ground motion is exceeded is determined from the assumption of a Poisson process ($P\left(A \geq\ A_0\right) = 1 - e^{-\lambda t}$). The reciprocal of this value defines the return period for the level of acceleration. For the example of the 10 % probability of being exceeded in 50 years, this equates to a 475 year return period. This assumption underpins time-independent seismic hazard analysis regardless of whether the Cornell-McGuire method or the Monte Carlo method is used. It is because of this that the earthquake catalogue input into the Monte Carlo programs used here is purged of aftershocks using the Reasenberg (1985) method.

The question to address is whether the hazard contribution from aftershocks is sufficiently trivial as to justify this assumption. Before considering this prospect, some clarification is needed as to what relevance aftershocks may have to engineering design. The standard requirements of a seismic hazard analysis are to identify the level of ground motion that has a P % probability of being exceeded in T years. The values attached to P and T depend on the structure being designed. This level of hazard is irrespective of whether the ground motion comes from the mainshock or an aftershock. In reality, aftershocks are of real significance as they further damage structures that may have already been weakened by the mainshock and previous seismicity. This can also have implications for emergency response following damaging earthquakes. There has been a substantial move toward real-time probabilistic aftershock modelling and hazard analysis in recent years (Gerstenberger *et al.*, 2004, 2005).

When moving beyond the assumption of stationary Poissonian seismicity, the reciprocity of return period and probability of a level of ground motion being exceeded is no longer valid. Beauval *et al.* (2006a) argue that a synthetic catalogue of length T_{SYN} can only be considered as N realisations of T years of seismicity, where $T_{SYN} = N \times T$. In doing so, they demonstrate that the inclusion of aftershocks in a Monte Carlo simulation of seismic hazard contributes to less than 5 % of the overall hazard. Although this contribution to seismic hazard may be small, between 20 % and 30 % of the earthquakes included in the synthetic catalogues were identified as aftershocks by the Reasenberg (1985) algorithm. It is apparent that although the impact of aftershocks on overall seismic hazard is small, there is merit in moving beyond the assumption of time-independent seismicity. The Monte Carlo method will therefore be used to incorporate aftershocks into the seismic hazard analysis.

7.6.2 The choice of aftershock simulation model

Across the many applications of statistical seismicity, there are a range of aftershock models in use. Two models that are in wide application are the Epidemic Type Aftershock Sequence (ETAS) model (Ogata, 1988, 1993) and the Short-Term Earthquake Probability (STEP) model (Reasenberg and Jones, 1989). Both models operate using two fundamental assumptions:

1. The aftershock process is a non-stationary Poisson process obeying the modified Omori law (Utsu, 1961):

$$\lambda_\theta = \frac{K}{(t+c)^p} \qquad \text{where} \qquad \theta = (K, c, p) \tag{7.6}$$

where t is the time (usually in days) after the mainshock and K, c and p are parameters to be estimated.

2. The aftershock magnitudes follow the Gutenberg and Richter (1944) relationship:

$$\log_{10} N_c(M) = a - bM \tag{7.7}$$

The STEP model is the simpler of the two aftershock simulators, which simply expresses rate of aftershocks (λ_θ) following a mainshock of magnitude (M_m) as a combination of the two assumptions:

$$\lambda_\theta(t, M) = 10^{a+b(M_m-M)}(t+c)^{-p} \tag{7.8}$$

The ETAS model is more complex and recognises that aftershocks are themselves capable of generating second and third order aftershocks. Furthermore, it also recognises that an aftershock sequence will be superimposed on top of the back ground seismicity, not in place of it. Therefore an additional parameter μ is included to model stationary Poissonian background seismicity. The ETAS model therefore depends on five parameters. The result is that the rate of aftershocks, with magnitude greater than the cut-off magnitude (M_{CO}), following an event of magnitude M_i at time t_i is given by:

$$\lambda_\theta(t) = \mu + \sum_{t_i<t} \frac{K_0 e^{\alpha(M_i-M_{co})}}{(t-t_i+c)^p} \tag{7.9}$$

Estimation of the parameters of the two models is done using a maximum likelihood method. For the STEP model, the log likelihood function to be maximised is given by (Ogata, 1983):

$$\begin{aligned}\ell(a,b,p,c) &= N[a+b(M_m-M_{co})]\ln(10) - \cdots \\ &\cdots - p\sum_{i=1}^{N}\ln(t_i+c) - 10^{a+b(M,-M_{\min})}\int_S^T (t+c)^{-p}\,dt\end{aligned} \tag{7.10}$$

where t_i and N are, respectively, the occurrence times and the total number of shocks in each sequence within the time range S to T (Gasperini and Lolli, 2006). The integral term is evaluated numerically. For the ETAS model, the log likelihood function is given by (Ogata, 1983, 1993):

$$\ell(\theta) = \sum_{i=1}^{N}\ln\lambda_\theta(t_i) - \int_S^T \lambda_\theta(t)\,dt \qquad \text{where} \qquad \theta = (\mu, K_0, \alpha, c, p) \tag{7.11}$$

The choice of which of the two models to use is difficult. Comparative fits of different aftershock models to observed aftershock sequences across the globe tend to suggest that the ETAS model captures observed aftershock sequences well, mainly because of the inclusion of secondary and tertiary aftershocks. However, the application of ETAS to probabilistic seismic hazard has not yet produced obviously successful results. One of

the reasons for this is the cutoff magnitude being considered. Maximum likelihood fits of the ETAS model to observed aftershock sequences (Ogata, 1993; Guo and Ogata, 1997; Liu and Ma, 2006; Beauval *et al.*, 2006a) generally assume lower cut-off magnitudes than the minimum magnitude being considered in the seismic hazard analysis here.

An implementation of the maximum likelihood method of calculating the ETAS parameters can be found in the IASPEI software library (Healy *et al.*, 1997). To determine the ETAS parameters, a search was conducted for aftershock sequences in the extended Burton *et al.* (2004a) catalogue. The AFTERAN declustering algorithm of Musson (1999b) was applied to the catalogue to identify aftershock sequences. AFTERAN was used for this purpose instead of the Reasenberg (1985) algorithm due to its larger spatial window. The aftershock sequences were then cut-off below the minimum magnitude used in the seismic hazard analysis (M_W 5.2). This last step removes a large number of aftershocks from each sequence, to the extent that only a few events contained enough aftershocks to estimate the parameters. Four aftershock sequences were input into the program: the 22 January 1912 Cephalonia event (M_W 6.3), 7 August 1915 Cephalonia event (M_W 6.5), 26 September 1932 Chalkidiki peninsula event (M_W 7.0) and the 12 August 1953 Cephalonia event (M_W 7.1). Despite the large number of damaging earthquakes recorded in the Aegean throughout the 20th century, these are the only events from which a sufficient number of aftershocks with $M_W \geq 5.2$ can be used to determine ETAS parameters. When these sequences were input into the program, the optimisation procedure failed to stabilise. The implementation in the IASPEI library uses a Davidon-Fletcher-Powell gradient descent method (Davidon, 1991) to maximise the log-likelihood function. With so few events in each sequence the output parameters varied considerably depending on the initial estimates in the descent algorithm.

Given the high cut-off magnitude it was not possible to estimate the ETAS parameters using the maximum likelihood method. A similar approach was therefore attempted using the likelihood estimate of the STEP method (equation 10). The same four aftershock sequences were used, this time optimising four parameters instead of five. Initially, maximisation of the log-likelihood was attempted using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method of unconstrained nonlinear optimisation. Parameter estimates were more stable than they has been for the ETAS model, but still displayed some variability.

Non-linear optimisation using quasi-Newton methods clearly illustrates how the small number of events in the aftershock sequence means that the maximum likelihood method cannot readily stabilise at a solution. A continuous-parameter genetic algorithm is therefore employed to identify the optimum set of parameters for minimising the negative log-likelihood function (10). A fixed population of 500 chromosomes (parameter estimates) is iterated for 500 generations. Rank weighting selection is applied with a 5 % mutation rate (Haupt and Haupt, 2004). A hybrid genetic algorithm is also compared alongside the conventional genetic algorithm. In the hybrid algorithm a single iteration

of the BFGS method is implemented as a heuristic to speed up convergence. The convergence towards the global optimum for the two algorithms is compared in Figure 7.29. This figure clearly shows that the genetic algorithm with the BFGS heuristic reaches is optimum sooner; whilst for the 1932 Chalkidiki event the non-hybrid genetic algorithm does not reach an optimum even after 500 generations.

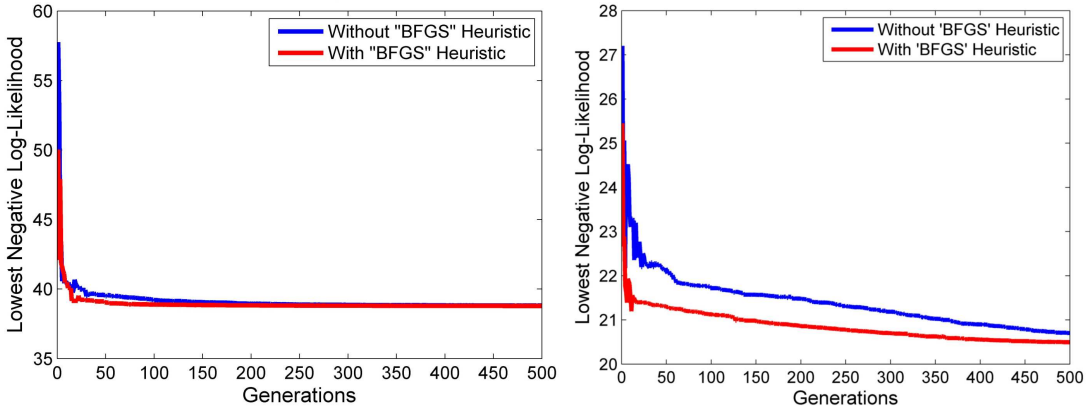

Figure 7.29: Convergence of fit toward the minimum negative log-likelihood to aftershock sequences from the a) 1953 Cephalonia event, b) 1932 Chalkidiki event

The 1932 and 1953 events contain the largest number of earthquakes. Estimates of p and c for the 1932 event are 0.9227 and 0.0325, respectively. For the 1953 Cephalonia event they are 0.7925 and 0.0475 for the p and c parameters respectively. The largest number of aftershocks is found in the 1953 Cephalonia sequence, which contains only 52 aftershocks. Although the genetic algorithms find stable solutions the magnitude parameters are influenced very strongly by the moderate-sized aftershocks, which can vary enormously in an aftershock sequence. The mainshock magnitudes are all in the range $6 \leq M_W \leq 7.1$. For a typical mainshock-aftershock sequence, Bath's law indicates the largest aftershock is usually around 1.2 ± 0.3 magnitude units smaller than the mainshock (Richter, 1958; Bath, 1965; Shcherbakov *et al.*, 2005). In all four of the sequences tested here the difference between the mainshock and largest aftershock was considerably smaller. Whether this is a particular property of the seismotectonic region in question (recalling that three of the four events are from the Ionian islands) is not clear. This deviation from Bath's law could also be coincidental or due to anomalous energy release from the mainshock in the region. Despite the endeavour, given the limited data set of aftershock sequences it is not possible to simultaneously optimise all four STEP parameters or all five ETAS parameters.

7.6.3 Aftershock Simulation in the Monte Carlo model

In lieu of empirically derived STEP parameters from the Aegean, the "generic California" model is adopted for the purposes of aftershock simulation. This model assumes the following values: a = -1.67, b = 0.91, p = 1.08 and c = 0.05. It refers to the median values determined from 62 sequences of aftershocks from California earthquakes (Reasenberg

and Jones, 1989). The c value is in good agreement with those observed in the Cephalonia and Chalkidiki aftershock sequences. The p-value is higher than those found in Greece, although the p-values of approximately 0.8 and 0.9 from the two Aegean earthquake sequences are within 1 standard deviation of the range of California aftershock sequences (Reasenberg and Jones, 1989).

In the aftershock simulator, Bath's law is assumed: $M_{AFT}^{MAX} \approx M_{MAIN} - \Delta M$ where $\Delta M = 1.2 \pm 0.3$ (Shcherbakov *et al.*, 2005). This serves two purposes. Firstly, it provides an upper bound estimate to the synthetic aftershock magnitudes. Secondly, it reduces the number of earthquakes from which aftershock simulation is attempted, effectively excluding aftershock simulation for magnitudes less than approximately M_W 6.0. Synthetic aftershock magnitudes are simulated using the truncated Gutenberg-Richter relation (see previous chapters) with a lower bound corresponding to M_{MIN} and an upper bound determined using Bath's law, with Gaussian scatter to allow for uncertainty in M_{AFT}^{MAX}.

The spatial aftershock distribution is simulated utilising some of the key assumptions found in the Real-time Earthquake Likelihood Model that has recently been used to forecast aftershocks following large earthquake in California (Gerstenberger *et al.*, 2004, 2005). Using the synthetic fault parameters (strike and rake) determined according to the method described in section 7.6. Assuming the Wells and Coppersmith (1994) empirical relation between subsurface rupture length (RLD) and M_W, a synthetic rupture is defined. Aftershocks are then simulated within a 0.5 $\times$ RLD radius of the synthetic rupture. The productivity of aftershocks is tapered, decreasing with distance from the rupture (r) such that the density decays with r^2. Although the ETAS model is not implemented, for larger earthquakes where secondary aftershocks may be greater than M_{MIN}, cascading sequences can be simulated. In such circumstances this is done by simply repeating the aftershock simulation for the larger aftershocks. Given the assumption of Bath's law and M_{MIN} = 5.2, secondary aftershocks are highly unlikely to be simulated for mainshocks less than approximately M_W 7.

Figure 7.30 shows a comparison of the seismic hazard maps for synthetic catalogues with aftershocks (b and d), and synthetic catalogues without aftershocks (a and c). It is once again apparent that despite the increase in the overall number of events (approximately 30 %) in the earthquake catalogue, the impact on seismic hazard is not especially great. This impact on seismic hazard (expressed as a percentage increase in PGA with a 10 % probability of not being exceeded in 50 years) is shown in Figure 7.31. Overall there is a slight increase in hazard at many sites, but on the order of no more than 5 %. Some areas show a slightly greater increase in hazard, whilst some other sites show a slight decrease. A proportion of this change in hazard may be attributable to the intrinsic noise of the Monte Carlo process. This may be the case in low hazard regions where a PGA increase of only a couple of $cm\ s^{-2}$ may represent a significant percentage increase in hazard. Although the impact on hazard is clearly spatially variable, there is no obvious

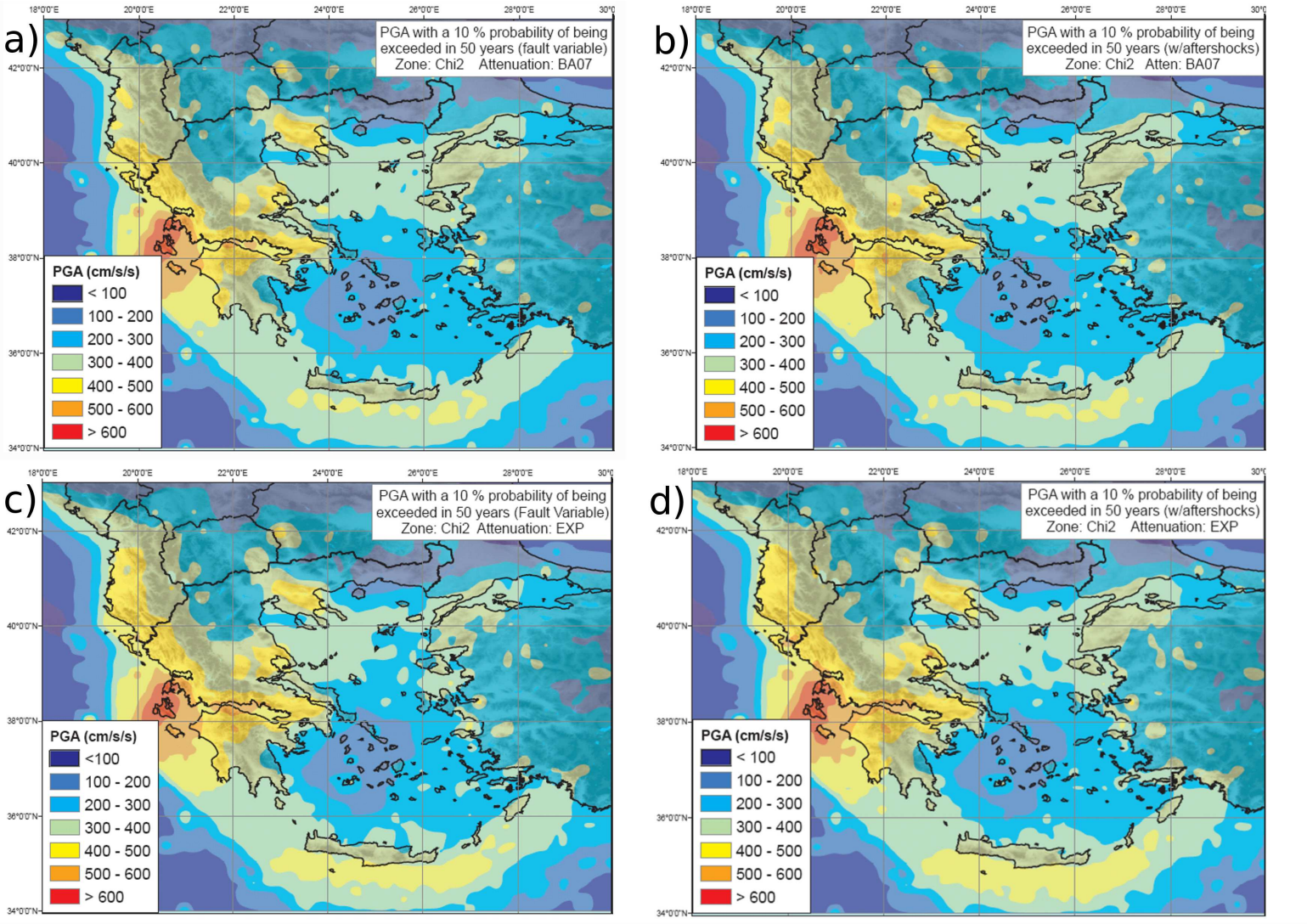

Figure 7.30: PGA with a 10 % probability of being exceeded in 50 years. a) χ^2 zone model, BA07 attenuation model (fault variable), b) as a) with aftershocks, c) χ^2 zone model, EXP attenuation model (fault variable), d) as c) with aftershocks.

spatial trend in the areas experiencing the greatest increase in hazard.

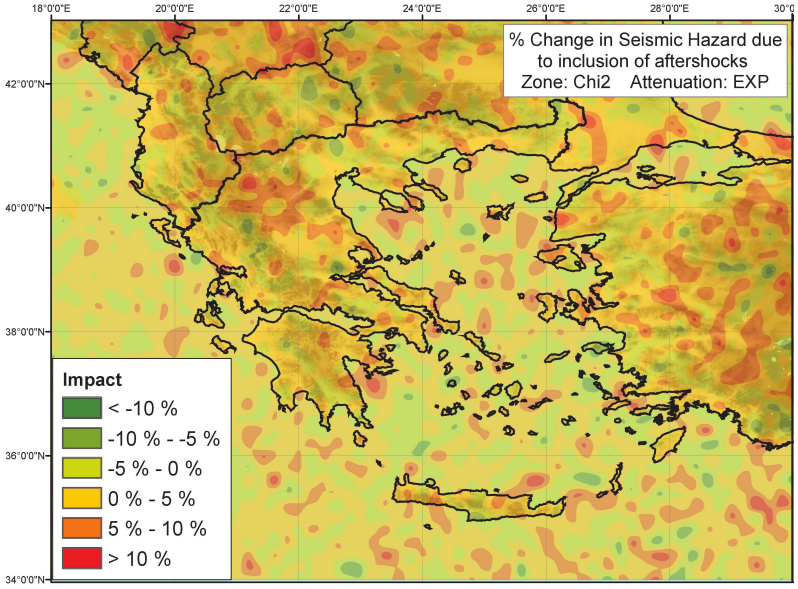

Figure 7.31: Increase in hazard (with a 10 % probability of being exceeded in 50 years) owing to the inclusion of aftershocks.

The inclusion of aftershocks illustrates how the Monte Carlo method can be extended beyond the assumption of stationary Poissonian seismicity. The inclusion of an aftershock simulator is a relatively simple approach to incorporating elements of time-dependent earthquake behaviour into the seismic hazard analysis. For a more detailed seismic hazard analysis, or at least a more localised analysis, more complex assumptions regarding the nature of time-dependence can be assumed. A lower minimum magnitude may allow for accurate determination of ETAS parameters, which was not achievable here. Models of self-exciting behaviour could also be incorporated into the analysis to consider foreshock behaviour. Given the small impact that inclusion of aftershocks have had on the overall seismic hazard level, the inclusion of foreshocks may be trivial.

The applications of time-dependent seismic hazard analysis are perhaps somewhat detached from the context of seismic hazard considered in this work. Mostly, time-dependent hazard analysis produces likelihoods of an earthquake magnitude, or level of ground motion, being exceeded in an area in the next T years. The Monte Carlo method is an effective tool for this particular application as it can effectively produce multiple realisations of the next T years of seismicity to analyse hazard (Beauval *et al.*, 2006a). The return periods being considered in this analysis are substantially longer than the brief durations often seen in time-dependent seismic hazard analysis. Given many of the non-linear aspects of earthquake behaviour it is unrealistic to expect that any method of time-dependent hazard analysis could give reasonable forecasts of earthquake behaviour for the 475 year or 975 year return period considered here. To reconcile, more completely, the Monte Carlo method of time-dependent seismic hazard analysis shown here with more practical requirements of such analyses, many more complex aspects of earthquake behaviour need to be taken into consideration. Some statistical methods and models are

being developed that may allow for time-dependent earthquake behaviour to be more effectively incorporated into seismic hazard analysis. These include marked stress release models (Rotondi and Varini, 2006, 2007) and cellular automata approaches (Jiménez *et al.*, 2005; Jiménez and Posadas, 2006; Georgoudas *et al.*, 2007). Arguably one of the most pressing challenges in this area is to be able to use Monte Carlo simulations of seismicity to characterise non-independent seismicity due to coulomb stress transfer (Stein *et al.*, 1997; King *et al.*, 1994). There are clearly many challenges in seismic hazard modelling for which the Monte Carlo method may be a powerful tool.

7.7 Conclusions

The material presented in this chapter demonstrates how the Monte Carlo method can be used to extend the seismic hazard analysis beyond the "single model" approach of the previous chapter. There are many modifications that have been made to the process that help provide a better estimate of the seismic hazard at a site, taking into account uncertainties and more realistic models of seismicity. The aim set out at the beginning of the chapter was to converge toward a single map or hazard analysis that could be considered a best guide to seismic hazard in the Aegean. The proviso to this being that for a valid hazard map, the methodology must be consistent across the region in question. The extent to which this aim has been accomplished needs to be assessed.

The modifications made to the standard Monte Carlo PSHA procedure in this chapter are summarised as follows:

1. Incorporation of epistemic uncertainty by integrating several competing (and appropriately weighted) models into the seismic hazard analysis.
2. Assessment of the impact of reducing M_{MIN} on the seismic hazard analysis, within the magnitude range for which attenuation models may be considered reasonably applicable.
3. Incorporation of intermediate-depth earthquakes into the Monte Carlo simulations.
4. Use of a global V_{S30} data set to assess the impact of site condition on the hazard maps.
5. Incorporation of simple zone-dependent classifications of fault type onto the hazard analysis, consistent with the required inputs of the attenuation models.
6. Incorporation of non-Poissonian events into the seismic hazard analysis by way of a simple "STEP" based aftershock simulator.

Whilst it may be judged that each of these modifications does affect the seismic hazard analysis, not all do so to the extent that they can be judged worthy of their additional

increase in model complexity. This is especially important when the uncertainties associated with these modifications are highly significant. The two obvious modifications for which this is the case is the inclusion of intermediate-depth earthquakes and of aftershocks. In both examples the introduction of more complex models did not result in a significant increase in hazard beyond the inherent variability of the Monte Carlo procedure. In the case of the former, the import of a non-Aegean attenuation model and the inconsistencies in the characterisation of intermediate-depth seismic source zones introduced errors that could not be readily quantified. In the latter, the difficulty in constraining the "STEP" model parameters, and the simplicity of some of the other assumptions in the aftershock simulator, meant that the aftershock process could only be modelled in a crude fashion. That neither the aftershocks nor the intermediate depth earthquakes produced a significant increase in hazard would suggest that they should not be included in the hazard analysis at this point. As these models become better constrained in future, this judgement could be reappraised.

The decrease in M_{MIN} tested in section 7.3 largely bore out the observations of previous studies into the sensitivity of this parameter. Hazard at shorter return periods is slightly higher when M_{MIN} is lowered, but this difference disappears with increasing return period. The impact of M_{MIN} is worth exploring in a hazard analysis, especially in a site-specific study. Depending on the location and the return period of interest for the structure under consideration, it may or may not be necessary to incorporate this parameter into the epistemic uncertainty analysis. The M_{MIN} of M_W 5.2 used in this study was robust, and consistent with the completeness of the catalogue, the lower bound magnitude of the attenuation relations used and the lower bound magnitude that may present a hazard to a site or structure. Decreasing M_{MIN} to M_W 4.8 or lower introduces bias from extrapolation of attenuation relations beyond their applicable limits (the exceptions being Bm07 and DT07). This simple sensitivity analysis would suggest that the M_{MIN} of M_W 5.2 selected in this hazard analysis is appropriate given the input models and data.

The modifications that show a clearer impact on hazard are the inclusion of site characterisation (by way of a global V_{S30} data set) and fault type. These models are generally simple and consistent with the zone and attenuation models used throughout this analysis. There are obvious questions that can be raised regarding both the use of V_{S30} as a site characterisation term, and the validity of the topographic approach to estimation of this term. Since the impact is in the strong motion attenuation, and three of the four attenuation models express site as NEHRP class rather than explicit V_{S30}, this data is still useful. The impact on hazard is more obvious and more realistic, illustrating how the geography of a region can affect the seismic hazard, even if the input models for the hazard remain the same. This has real implications for hazard mapping and it is fortunate that this database has been made available during this work.

The incorporation of fault type is also done by a simple approximation of the tectonics, which is consistent with the approximations of fault type made in the attenuation

models. Again, this is blurring the line between site-specific hazard analysis and hazard mapping, but the results do have important implications for mapping purposes. There is a clear increase in hazard along the Hellenic and Adriatic coastlines when thrust faulting is included in the model. Although the delineations of zones of different fault types are crude, this modification is necessary. Most attenuation models developed in the last ten years have included a faulting parameter term. This removes the variability in fault type from the aleatory uncertainty captured by the attenuation model, and places it into the epistemic uncertainty category. Assuming, not unreasonably, that the attenuation models were created with site-specific hazard analysis rather than hazard mapping in mind, this move is understandable. When mapping seismic hazard, to incorporate fault type a spatially consistent approach is needed. The K-means source models offer one way to achieve this, but this chapter has also shown that simple judgements regarding the predominant type of faulting in pre-defined source zones can also be used. It is therefore important this modification be kept in the hazard analysis.

Arguably the most significant modification to the Monte Carlo procedure made in this chapter is the MCMAEU method for analysis of epistemic uncertainty. It has illustrated how the diversity of hazard maps shown in the previous chapter can be incorporated into the uncertainty analysis to produce a single map. The similarity of the results using MCMAEU and a traditional logic tree approach demonstrate its consistency with existing seismic hazard analysis procedures. The principal shortcoming of this method, and that of the logic tree, is the issue of selecting appropriate weighting schemes for the epistemic uncertainty analysis. As a matter of personal judgment, the best recommended scheme is the χ^2 zone model weighting and EXP attenuation model weighting. These broadly represent the best fit of the respective models to the input data used. The similarity of the hazard maps with the different schemes tested here, would suggest that a modest degree of perturbation in the weighting schemes does not necessarily have large impact on hazard. In response to the primary aim of this chapter, which was convergence toward a single hazard map or analysis, the MCMAEU method with χ^2 based weighting for the zone model and EXP weighting for the attenuation model is most appropriate for the production of a seismic hazard map in the Aegean. This map is shown in the next chapter.

Chapter 8

Conclusions and Future Directions

8.1 Summary of the Results of this Thesis

The material presented in this thesis details the necessary inputs and concepts for a seismic hazard analysis in the Aegean region. A clear progression is described, from consideration of statistical earthquake behaviour and recurrence, delineation of source models (here using a novel algorithmic technique), selection of attenuation models and finally simulation of synthetic catalogues. Presented in chapters 6 and 7 are the results of the seismic hazard analyses implemented using the Monte Carlo method, with several new extensions.

Initial analysis of the input earthquake catalogue (Chapter 2) identified several key parameters to assist in the constraint of probabilistic seismic hazard in the Aegean region. Analysis of maximum magnitude via several different methods indicates that M_W 8.6 may be an appropriate regional maximum magnitude. Statistical methods to analyse local variation in M_{MAX} result in estimates with greater uncertainties, in most cases substantially lower than regional M_{MAX}. Completeness magnitude (M_C) and its temporal variation are also considered. Four eras of completeness are clearly identified: historical (pre-1900 A.D. - $M_C \approx M_W, 6.8 - 7.0$), early instrumental (1900 A.D. - 1963 A.D. - $M_C \approx M_W$ 5.2), ISC instrumental (1964 A.D. - 1998 A.D. - $M_C \approx M_W$ 4.8) and NOA instrumental (1999 A.D. - 2005 A.D. - $M_C \leq M_W$ 4.0). Given the temporal variation in completeness magnitude, and taking into consideration the demands of seismic hazard analysis for engineering design, a minimum magnitude of M_W 5.2 is adopted in the seismic hazard analysis. Comparison of the characteristic earthquake recurrence model and the doubly-truncated Gutenberg-Richter earthquake model is made. There appears to be insufficient justification, both theoretically and empirically, to prefer the characteristic earthquake model for use in the Aegean; hence the truncated Gutenberg-Richter model is used here.

The delineation of shallow seismic source zones for use in seismic hazard mapping is a particularly complex issue in an area as tectonically diverse as the Aegean. Existing models are discussed and compared in chapter 3. Using current interpretations of seismotectonics in the Aegean region, a new source model is introduced here. This model identifies 11 shallow uniform zones, and two intermediate depth zones. This is done using a traditional approach, whereby sources are delineated by judgement, with regard to a wide range of seismotectonic and geophysical data. Outside of the uniform zones seismic source delineation is guided by the observed distribution of seismicity.

Recognising the subjectivity, opacity and irreproducibility of the "expert judgement" approach to source delineation, a novel algorithmic approach is implemented in chapter 4. K-means cluster analysis has been applied to the spatial distribution of hypocentres and fault ruptures, for the purposes of zone delineation. Common clustering metrics were initially used to identify the optimum number of partitions. These were found to be inconsistent when different subsets of the data are compared. A further step was taken whereby partitions were assessed not on their clustering quality, but on their ability to replicate observed hazard when translated into seismic source zones. Although some variability still exists, source models with approximately 27 to 30 zones emerge as the best at reproducing seismic hazard. The K = 27 and K = 29 partitions of fault ruptures in the Aegean are used in the subsequent hazard analyses.

The topic of strong ground motion attenuation and its uncertainties, both epistemic and aleatory, is addressed in chapter 5. This illustrates the decision making process required to select appropriate attenuation models for the Aegean region. A qualitative assessment of attenuation models resulted in the identification of six candidate models for the attenuation of PGA from shallow earthquakes in the Aegean: Am96, Sk03, Am05, DT07, Bm07 and BA07. Quantitative maximum likelihood assessment of the fit of these models to observed strong motion data helped refine this list (rejecting Am96 and Sk03), and provides a basis for weighting the remaining relations in an epistemic uncertainty analysis. The Am05, DT07, Bm07 and BA07 relations were identified as applicable candidate relations. The Am05 relation appears to produce unrealistically high ground motions in the seismic hazard analysis, whilst DT07 produces lower hazard estimates. For PGA and spectral acceleration attenuation, the BA07 relation appears to be the most appropriate. The Bm07 relation produces similar hazard estimates of a similar level as those produced using BA07, but the heteroscedastic modelling of aleatory variability, and the limited spectral range of the Bm07 model, result in BA07 being preferred.

Chapter 6 presents the results of the seismic hazard analyses comparing each of the different models separately. This demonstrates the impact that different combinations of source and attenuation model have on both spatial and site-specific seismic hazard analysis. The spatial distribution appears to be strongly controlled by the source model, whilst the actual value of hazard at a site differs considerably for some attenuation models. There is reasonable agreement in the PGA hazard curves at different sites across the

Aegean when the BA07 and Bm07 relations are used. Larger hazard estimates are produced when the Am05 attenuation relation is used, at some locations in excess of 1g for the 475-year return period. This substantial contrast from the behaviour of other models would suggest that the Am05 relation should not be used in future seismic hazard analyses in this region.

Disaggregation of earthquakes contributing to hazard at a site reveals some of the problems underlying these hazard analyses. At many sites it is low-magnitude, near-field earthquakes, with aleatory variability sampled above 1σ that produce the level of acceleration for a specified return period. This observation is also borne out for seismic hazard analyses in California using the Cornell-McGuire method (McGuire, 1995). This is contrary to the types of earthquakes that would most intuitively pose the greatest risk to the site: larger earthquakes at near to moderate distances with strong motion sampled close to the median. This illustrates a problem that appears to be emerging as ground motion aleatory variability is more widely incorporated into seismic hazard analysis. It is known that uncertainty on peak ground motions varies according to a lognormal distribution. Within the $\pm 3\sigma$ band considered here, this distribution has been shown to be a valid supposition (Bommer *et al.*, 2004; Strasser *et al.*, 2008). Under the assumption of a uniform source zone, where the target site lies within the source zone, over long duration synthetic catalogues a large number of earthquakes will be simulated within a short distance of the site. Assuming that the aleatory variability on each of these events is sampled randomly, and is independent of any other event, then it is expected that for earthquake catalogues of 10^7 to 10^8 years duration, many of the small near-field events will produce PGA well above the median forecast by an attenuation model. Whilst this will clearly influence the PGA hazard, it does not mean that small near field events should form the basis for engineering design, especially when stronger events at greater distances are also capable of generating the same PGA at a site, even if the smaller events are more abundant! This is a difficulty currently inherent in this methodology.

The extensions to the Monte Carlo PSHA process shown in chapter 7 are novel developments designed with the intention of producing a "recommended" hazard map. Here, epistemic uncertainty is incorporated into the Monte Carlo procedure, which means several competing models can be combined into the same hazard analysis. Unlike the Logic Tree procedure, no decision need be made as to whether the mean or median hazard is preferred, although the hazard is slightly higher than either of the mean or median. Several other features of seismicity, often not addressed in seismic hazard analyses, have also been tested. The inclusion of seismic hazard from intermediate-depth earthquakes has little impact on the overall regional seismic hazard analysis. Likewise the inclusion of aftershocks did not result in a significant change in the overall seismic hazard at the return periods being considered here. In both of these cases the increase in complexity, and the difficulty in accurately constraining particular facets of the models (seismic source and attenuation for deep earthquakes; rate and time distribution of aftershocks), are not necessarily justified given the insignificant change in overall hazard. In contrast,

the inclusion of simple faulting terms and site characteristics has an appreciable impact on seismic hazard.

8.2 A "Recommended" Seismic Hazard Map for the Aegean Region"

All of the maps shown in chapters 6 and 7 depict the seismic hazard in the Aegean, albeit making specific assumptions in the modelling process. This suite of maps is useful in demonstrating the degree of variability in the seismic hazard when different models of source and attenuation are considered. This alone may constitute a comprehensive seismic hazard analysis, and whilst guidance as to which models (or model combinations) may be preferred, a "recommended" seismic hazard map has not yet been shown. Taking into account the results of chapters 6 and 7, the "recommended" seismic hazard map is shown in Figure 8.1.

The "recommended" seismic hazard map, showing the PGA (in cm s^{-2}) with a 10 % probability of being exceeded in 50 years, is constructed using the following conditions:

1. PSHA determined using the Monte Carlo Method
2. Epistemic Uncertainty integrated using MCMAEU with the χ^2 zone weighting scheme (incorporating the PP2000, PZ1990, WT2006, K27 and K29 source models) and EXP attenuation weighting scheme (incorporating the Am05, DT07, Bm07 and BA07 attenuation models).
3. Site condition is incorporated via NEHRP site class, using topographically derived V_{s30} (Wald and Allen, 2007) as an indicator
4. Fault type is incorporated into the analysis and varies on a zone by zone basis.
5. Seismicity contributing to hazard occurs at depths no greater than 60 km.
6. Non-Poissonian events (aftershocks) are not included in the simulation
7. $M_{MIN} = M_W\, 5.2$
8. PGA refers to the geometric mean of the two horizontal components of strong motion

The use of inverted commas in the term "recommended" indicate that this map represents the author's personal judgement as to which model combination and preliminary conditions are preferred in this seismic hazard analysis. Epistemic uncertainty is incorporated into the analysis using the MCMAEU method, to avoid making a judgement as to whether the mean or median level of hazard should be used.

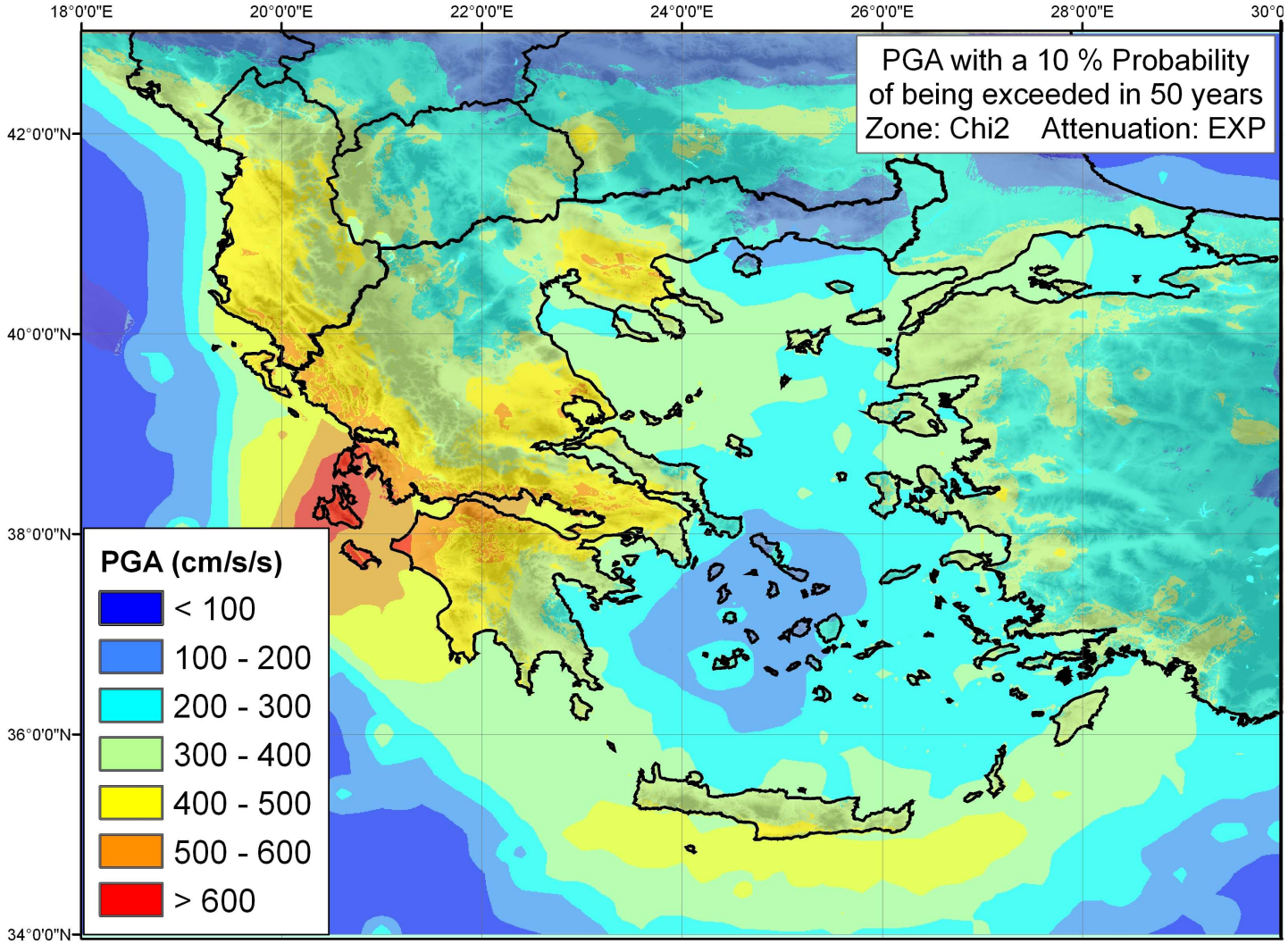

Figure 8.1: A preferred seismic hazard map. PGA (in $cm\ s^{-2}$) with a 10 % probability of being exceeded in 50 years in the Aegean region. See text for details of the model used.

8.3 Implications of the Research

The results of this thesis can be divided into two contexts. The first is the procedural aspect of seismic hazard analysis. The second is the seismic hazard in the Aegean region.

For the procedural context, this research has developed new techniques, and expanded the application of existing techniques to new areas. Whilst the Monte Carlo approach to PSHA is not novel, several extensions to the method have been made that do develop it in novel ways (both geographically and conceptually). Possibly the most significant new approach is the delineation of seismic source zones using the K-means methodology. It is shown to be applicable to the Aegean region, even if several issues remain an area of user judgement. This technique can be applied anywhere in the world where a homogenous earthquake catalogue can be constructed. The line K-means variant may be more limited to areas where there is existing quantitative information regarding the location and size of active faults. The K-means cluster analysis approach has already been applied to Indonesia (Burton *et al.*, 2008), where it has helped delineate seismic sources on scales much smaller than those considered in previous seismic hazard analyses for the region. It is hoped that this work, and the accompanying journal paper (Weatherill and Burton, 2009), will form a basis from which the K-means methodology for source zone delineation can be developed.

The extensions to the Monte Carlo technique covered in chapter 7 (e.g. MCMAEU, site condition, fault condition, aftershock simulation etc.) move the Monte Carlo method beyond the basic approaches found in previous studies (Ebel and Kafka, 1999; Musson, 1999b; Giardini *et al.*, 2004). In the case of site condition, this development is in no small part due to the publication and dissemination of the global V_{S30} database (Wald and Allen, 2007). However, the extension of the Monte Carlo method to incorporate epistemic uncertainty represents a hitherto undeveloped direction for the procedure.

In terms of the implications for seismic hazard analysis in the Aegean region specifically, this work represents a new progression. The development of new seismic source models, and their ultimate assimilation into a single hazard analysis procedure, begins to illustrate the significance of the source model in the context of epistemic uncertainty in seismic hazard in the Aegean. Equally, the review and selection of attenuation models (including NGA models published during the period of this work) appropriate to the Aegean can be used as guidance for the choice of attenuation model to use in future seismic hazard analysis. The seismic hazard maps can be used to form the basis for building codes in Greece, Albania and FYROM, as they are compatible with existing methodologies of seismic hazard analysis.

8.4 Targets for future research

Throughout the course of this research several issues have emerged that are important targets for future research into seismic hazard in the Aegean region. They can be expressed as four research aims:

8.4.1 Resolution of the debate regarding homoscedastic versus heteroscedastic aleatory variability in the attenuation relation

Although there are many differences between the attenuation models used in this work, one of the most significant is the difference between those that have a magnitude dependent aleatory variability term (Am05 and Bm07) and those that don't (DT07 and BA07). The increase in the sigma term at lower magnitudes has the effect of diminishing the influence of larger magnitude events in the seismic hazard analysis. A case for magnitude-dependent variability is made by Youngs *et al.* (1995), who ascertain that it is not merely an artefact of the empirical data distribution, nor the non-linearity of site response. This does not necessarily resolve the issue as there are still many facets of uncertainty in strong motion records due to anthropogenic factors, many of which are affected by earthquake magnitude. These include errors in estimation of magnitude, location and stress drop of small events compared to large events. Both homo- and heteroscedastic approaches are still widely used. It is hoped that as the global database of strong motion records expands new insights will emerge.

8.4.2 Improvement of strong ground motion attenuation models for intermediate depth and deep earthquakes

The paucity of strong motion data from intermediate depth earthquakes in the Aegean region remains a problem (Chapter 7). It may be assumed that over time both the number of such records from the Aegean, and the number of deep earthquake attenuation models (global), will increase. This expansion should eventually enable maximum likelihood estimation of fit of existing attenuation models to Greek data. The ultimate aim should be to develop attenuation models that are specific to intermediate depth Greek earthquakes. Failing this, a greater contribution of Greek records to global sets of strong motion data from deep events should also be seen as an objective.

8.4.3 Comparison of strong motion attenuation variability between mainshocks and foreshocks/aftershocks

The separation of inter- and intra-event variability that is made using the random effects regression procedure (Brillinger and Preisler, 1984) makes the fundamental assumption that each separate event is independent. Many of the strong motion data sets used for re-

gression contain mixtures of records from mainshocks and aftershocks. Simulating variability in the random manner ascribed here means that the σ value sampled for aftershocks is different from, and entirely independent of that of the main shocks. In reality it is likely that both mainshocks and aftershocks will share some common characteristics that will affect the attenuation of strong motion (e.g. directivity, faulting term etc), which may invalidate the assumption of independence between events. If Monte Carlo methods are to be extended to include aftershock simulation in the manner demonstrated here, it needs to be established whether, for the purposes of strong ground motion modelling, these can be treated as independent events. Were this to be an unsafe assumption, there may be a reduction in inter-event variability when considering mainshock-aftershock sequences. This could ultimately affect the overall seismic hazard at a site, and will have an impact on time-dependent forecasts of strong ground motion that are made following large earthquakes.

8.4.4 Quantitative evaluation of the fit of source models to a region

Evaluation of source models is an area of seismic hazard analysis that retains a lack of transparency and objectivity. Quantitative measures of fit such as those suggested in this thesis or by Musson (2004) may be useful, but they suffer from some shortcomings. Firstly, these measures implicitly assume that the distribution of seismicity in the observed catalogue is an accurate representation of the true distribution of seismicity over time (for the likely return periods being considered in the hazard analysis). Depending on the length of the observed catalogue this may be a reasonable assumption. The second problem is that this method can only fairly compare source models that are developed in the same way. This means that uniform zone source models cannot be compared with hybrid (uniform and non-uniform zone) models or with smoothed seismicity approaches to source zonation. The latter two approaches should be more strongly correlated with observed seismicity, since they are derived directly from the distribution of observed seismicity. They will therefore be afforded lower χ^2 values.

The seismic source model (and indeed the validity of any seismic hazard analysis) cannot be evaluated on the basis of hypocentral distribution or magnitude-frequency distribution exclusively. To appraise a seismic hazard analysis effectively, comparison must be made with the distribution of the observed parameter used in the seismic hazard analysis. A very general comparison between the hazard curve and observed occurrence of macroseismic intensities for Argostoli and Istanbul was made in chapter 6. Further research is needed that consider three timescales: 1) the use of strong motion records to evaluate the validity of the PSHA (in PGA) for short return periods (approximately 30 years depending on the location), this approach has already been started elsewhere in the world but with limited results (Ordaz and Reyes, 1999; Beauval *et al.*, 2008); 2) use of macroseismic intensity records to evaluate the seismic hazard in terms of intensity for longer return periods (40 to several hundred years); 3) the use of precariously balanced

rocks to evaluate hazard estimates over geological timescales (several thousand years). Only Beauval *et al.* (2008) has considered this approach for a selected site in the Aegean. To evaluate the source model effectively, these approaches need to be implemented over a wider spatial scale.

8.5 Final Comments

This thesis has demonstrated the methodology of seismic hazard analysis via Monte Carlo simulation and applied it to the Aegean region. The final result is a suite of maps that depict the seismic hazard; each constructed using different input models, and a final map that represents this author's judgement as to the most appropriate for the Aegean. The scale of epistemic uncertainty has been well-illustrated here. Considerable discussion regarding many of the most fundamental assumptions of seismic hazard analysis has been given throughout. Several novel techniques to refine the hazard analysis, with a focus on increasing the objectivity of the process, have been demonstrated, and their effectiveness assessed.

The hazard map(s) that arise from this thesis can be used to form the basis of seismic zonation for the implementation of Eurocode 8 in the Aegean. They may also be used for vulnerability studies in urban areas across the region. Site-specific data should not be extracted from these maps. However, the Monte Carlo methodology adopted here has been shown to be capable of undertaking such assessment, albeit a more localised approach would be needed for a robust site-specific analysis. This work is a progressive step toward improving the estimation of seismic hazard in the Aegean region, and identification of its uncertainties.

Bibliography

Abrahamson N. 2000. State of the practice of seismic hazard evaluation. In: *GEOEng 2000*. Australian Geomechanics Society: Melbourne, Australia, pp. 1–27.

Abrahamson N. 2006. Seismic hazard assessment: Problems with current practice and future developments. In: *First European Conference on Earthquake Engineering and Seismology*. ECEES: Geneva, Switzerland.

Abrahamson N, Silva WJ. 2007. Abrahamson and Silva NGA ground-motion relations for the geometric mean horizontal component of peak and spectral ground motion parameters. Peer report, Pacific Earthquake Engineering Research Center.

Abrahamson N, Youngs RR. 1992. A stable algorithm for regression analyses using the random effects model. *Bulletin of the Seismological Society of America* **82**(1): 505–510.

Abrahamson NA, Bommer JJ. 2005. Probability and uncertainty in seismic hazard analysis. *Earthquake Spectra* **21**(2): 603 – 607.

Aki K. 1965. Maximum likelihood estimate of b in the formula Log N = a - bm and its confidence limits. *Bullettin of the Earthquake Research Institute., Tokyo University* **43**: 237 – 238.

Al Abbasi JN, Fahmi KJ. 1991. GEMPAK: A Fortran-77 program for calculating Gumbel's first, third and mixture upper earthquake magnitude distributions employing maximum likelihood estimation. *Computers and Geosciences* **17**(2): 271 – 290.

Allen MP, Evans GT, Frenkel D, Mulder BM. 1993. Hard Convex Body Fluids. *Advances in Chemical Physics* **84**: 1 – 166.

Allen TI, Wald DJ. 2007. Topographic slope as a proxy for seismic site-conditions (Vs30) and amplification around the globe,. Technical report, United States Geological Survey, Open-File Report 2007-1257.

Ambraseys NN. 1995. The prediction of peak ground acceleration in Europe. *Earthquake Engineering and Structural Dynamics* **24**: 467–490.

Ambraseys NN, Bommer JJ. 1991. Database of European strong ground-motion records. *European Earthquake Engineering* **2**: 18 – 37.

Ambraseys NN, Bommer JJ. 1992. The attenuation of ground accelerations in Europe. *Earthquake Engineering and Structural Dynamics* **20**: 1179 – 1202.

Ambraseys NN, Douglas J, Sarma SK, Smit PM. 2005a. Equations for the estimation of strong ground motions from shallow crustal earthquakes using data from Europe and the Middle East: Horizontal peak ground acceleration and spectral acceleration. *Bulletin of Earthquake Engineering* **3**: 1 – 53.

Ambraseys NN, Douglas J, Sarma SK, Smit PM. 2005b. Equations for the estimation of strong ground motions from shallow crustal earthquakes using data from Europe and the Middle east: Vertical peak ground acceleration and spectral acceleration. *Bulletin of Earthquake Engineering* **3**: 55 – 73.

Ambraseys NN, Jackson JA. 2000. Seismicity of the Sea of Marmara (Turkey) since 1500. *Geophysical Journal International* **141**: F1–F6.

Ambraseys NN, Simpson KA, Bommer JJ. 1996. Prediction of horizontal response spectra in Europe. *Earthquake Engineering and Structural Dynamics* **25**: 371–400.

Ambraseys NN, Smit P, Douglas J, Margaris B, Sigbj´ornsson R, Ólafsson S, Suhadolc P, Costa G. 2004. Internet-site for European strong motion data. *Bollettino di Geofisica Teorica ed Applicata* **45**(3): 113 – 129.

Anderson H, Jackson J. 1987. Active tectonics of the Adriatic region. *Geophysical Journal of the Royal Astronomical Society* **91**: 937–983.

Anderson JG, Brune JN. 1999. Probabilistic seismic hazard analysis without the ergodic assumption. *Seismological Research Letters* **70**(1): 19 – 28.

Anooshehpoor A, Brune JN, Zheng Y. 2004. Methodology for obtaining constraints on ground motion from precariously balanced rocks. *Bulletin of the Seismological Society of America* **94**(1): 285 – 303.

Ardeleanu L, Leydecker G, Bonjer KP, Busche H, Kaiser D, Schmitt T. 2005. Probabilistic seismic hazard map for Romania as a basis for a new building code. *Natural Hazards and Earth System Sciences* **5**(5): 679 – 684.

Arias A. 1970. A measure of earthquake intensity. In: *Seismic Design for Nuclear Power Plants*, Hansen R (ed), MIT Press: Cambridge, Massachusetts, pp. 438 – 483.

Armijo R, Lyon-Caen H, Papanastassiou D. 1992. East-west extension and Holocene normal fault scarps in the Hellenic arc. *Geology* **20**: 491–494.

Armijo R, Meyer B, Navarro S, King G, Barka A. 2002. Asymmetric slip partitioning in the Sea of Marmara pull-apart: A clue to propagation processes of the North Anatolian Fault. *Terra Nova* **14**: 80–86.

Atkinson GM, Boore DM. 2003. Empirical ground-motion relations for subduction-zone earthquakes and their application to Cascadia and other regions. *Bulletin of the Seismological Society of America* **93**(4): 1703 – 1729.

Atkinson GM, Boore DM. 2008. Erratum to Empirical ground-motion relations for subduction zone earthquakes and their application Cascadia and other regions. *Bulletin of the Seismological Society of America* **98**(5): 2567 – 2569.

Avallone A, Briole P, Agatza-Balodimou AM, Billiris H, Charade O, Mitsakaki C, Nercessian A, Papazissi K, Paradissis D, Veis G. 2004. Analysis of eleven years of deformation measure by GPS in the Corinth Rift Laboratory area. *Comptes Rendus Geosciences* **336**: 301–311.

Baker C, Hatzfield D, Lyon-Caen H, Papadimitriou E, Rigo A. 1997. Earthquake mechanisms of the Adriatic sea and Western Greece: implications for the oceanic subduction-continental collision transition. *Geophysical Journal International* **131**: 559–594.

Barka A. 1999. The 17 August 1999 Izmit Earthquake. *Science* **285**(5435): 1858 – 1859.

Barka A, Ayküz HS, Altunel E, Sunal G, Çakir Z, Dikbas A, Yerli B, Armijo R, Meyer B, De Chabalier JB, Rockwell T, Dolan JR, Hartleb R, Dawson T, Christofferson S, Tucker A, Fumal T, Langridge R, Stenner H, Lettis W, Bachhuber J, Page W. 2002. The surface rupture and slip distribution of the 17 August 1999 Izmit earthquake (M 7.4), North Anatolian Fault. *Bulletin of the Seismological Society of America* **92**(1): 43 – 60.

Bath M. 1965. Lateral inhomogeneities of the upper mantle. *Tectonophysics* **2**(6): 483 – 514.

Bazzurro P, Cornell CA. 1999. Disaggregation of seismic hazard. *Bulletin of the Seismological Society of America* **89**(2): 501–520.

Beauval C, Bard PY, Hainzl S, Gueguen P. 2008. Can strong-motion observations be used to constrain probabilistic seismic-hazard estimates? *Bulletin of the Seismological Society of America* **98**(2): 509 – 520.

Beauval C, Hainzl S, Scherbaum F. 2006a. Probabilistic seismic hazard estimation in low-seismicity regions considering non-Poissonian seismic occurrence. *Geophysical Journal International* **164**: 543–550.

Beauval C, Scotti O. 2004. Quantifying seinsitivities of PSHA for France to earthquake catalog uncertainties, truncation of ground-motion variability and magnitude limits. *Bulletin of the Seismological Society of America* **94**(5): 1579–1594.

Beauval C, Scotti O, Bonilla F. 2006b. The role of seismicity models in probabilistic seismic hazard estimation: comparison of a zoning and a smoothing approach. *Geophysical Journal International* **165**: 584–595.

Bender B. 1983. Maximum likelihood estimation of b values for magnitude grouped data. *Bulletin of the Seismological Society of America* **73**(3): 831–851.

Bender B, Campbell KW. 1989. A note on the selection of minimum magnitude for use in seismic hazard analysis. *Bulletin of the Seismological Society of America* **79**: 199 – 204.

Bender B, Perkins DM. 1987. SEISRISK III: A computer program for seismic hazard evaluation. Technical report, USGS.

Bendimerad F, Coburn A, Morro G, Johnson L, Rahnama M. 2000. Event report: Kocaeli, Turkey earthquake. Technical report, Risk Management Solutions, Inc.

Benetatos C, Kiratzi A, Papazachos C, Karakaisis G. 2004. Focal mechanisms of shallow and intermediate depth earthquakes along the Hellenic Arc. *Journal of Geodynamics* **37**: 253–296.

Bernard P, Lyon-Caen H, Briole P, Deschamps A, Boudin F, Makropoulos K, Papadimitriou P, Lemeille F, Patau G, Billiris H, Paradissis D, Papazissi K, Castarède H, Charade O, Nercessian A, Avallone A, Pacchiani F, Zahradnik J, Sacks S, Linde A. 2006. Seismicity, deformation and seismic hazard in the western rift of Corinth: New insights from the Corinth Rift Laboratory (CRL). *Tectonophysics* **426**: 7–30.

Bindi D, Luzi L, Pacor F, Franceschina G, Castro RR. 2006. Ground-motion predictions from empirical attenuation relationships versus recorded data: The case of the 1997-1998 Umbria-Marche, central Italy, strong-motion data set. *Bulletin of the Seismological Society of America* **96**(3): 984 – 1002.

Bohnhoff M, Harjes HP, Meier T. 2005. Deformation and stress regimes in the Hellenic subduction zone from focal mechanisms. *Journal of Seismology* **9**: 341–366.

Bommer JJ, Abrahamson NA. 2006. Why do modern probabilistic seismic hazard analyses often lead to increased hazard estimates. *Bulletin of the Seismological Society of America* **96**(6): 1967 – 1977.

Bommer JJ, Abrahamson NA, Strasser FO, Pecker A, Bard PY, Bungum H, Cotton F, Fäh D, Sabetta F, Scherbaum F, Studer J. 2004. The challenge of defining upper bounds on earthquake ground motions. *Seismological Research Letters* **75**(1): 82–94.

Bommer JJ, Douglas J, Strasser FO. 2003. Style-of-faulting in ground-motion prediction equations. *Bulletin of Earthquake Engineering* **1**: 171–203.

Bommer JJ, McQueen C, Salazar W, Scott S, Woo G. 1998. A case study of the spatial distribution of seismic hazard (El Salvador). *Natural Hazards* **18**: 145–166.

Bommer JJ, Scherbaum F, Bungam H, Cotton F, Sabetta F, Abrahamson NA. 2005. On the use of logic trees for ground-motion prediction equations in seismic-hazard analysis. *Bulletin of the Seismological Society of America* **95**(2): 377 – 389.

Bommer JJ, Stafford PJ, Alarcón JE, Akkar S. 2007. The influence of magnitude range on empirical ground motion. *Bulletin of the Seismological Society of America* **97**(6): 2152 – 2170.

Boore DM, Atkinson GA. 2007. Boore and Atkinson NGA ground-motion relations for the geometric mean horizontal component of peak and spectral ground motion parameters. Peer report, Pacific Earthquake Engineering Research Center.

Boore DM, Watson-Lamprey J, Abrahamson NA. 2006. Orientation-independent measures of ground motion. *Bulletin of the Seismological Society of America* **96**(4A): 1502 – 1511.

Boore D M amd Joyner WB, Fumal TE. 1997. Equations for estimating horizontal response spectra and peak acceleration from western North American earthquakes: A summary of recent work. *Seismological Research Letters* **68**(1): 128 – 153.

Bott MHP. 1993. Modelling the plate driving mechanism. *Journal of the Geological Society of London* **150**(5): 941–951.

Bradley PS, Fayyad UM. 1998. Refining initial points for K-means clustering. In: *Proceedings of the 15th International Conference on Machine Learning*, Shavlik J (ed). Morgan Kauffman: San Francisco, pp. 91 – 99.

Brillinger DR, Preisler HK. 1984. An exploratory analysis of the Joyner-Boore attenuation data. *Bulletin of the Seismological Society of America* **74**: 1441 – 1450.

Brune JN. 1996. Precariously balanced rocks and ground-motion maps for southern California. *Bulletin of the Seismological Society of America* **86**(1A): 43 – 54.

Burchfiel BC, King RW, Todosov A, Kotzev V, Durmurdzanov N, Serafimovski T, Nurce B. 2006. GPS results for Macedonia and its importance for the tectonics of the South Balkan extensional regime. *Tectonophysics* **413**: 239–248.

Burnham KP, Anderson DR. 2002. *Model selection and multi-model inference: A practical information-theoretic approach*. Springer-Verlag: New York, 2nd edn.

Burnham KP, Anderson DR. 2004. Understanding AIC and BIC in model selection. *Sociological Methods and Research* **33**: 261 – 304.

Burton PW. 1979. Seismic risk in southern Europe through to India examined using Gumbel's third distribution of extreme values. *Geophysical Journal of the Royal Astronomical Society* **59**: 249 – 280.

Burton PW, Weatherill G, Karnawati D, Pramumijoyo S. 2008. Seismic hazard assessment and zoning in Java: New and alternative probabilistic assessment models. In: *Proceedings of the International Conference on Earthquake Engineering and Mitigation*. Jakarta, 14 - 15 April 2008.

Burton PW, Xu Y, Changyuan Q. 2001. Earthquake strong motion attenuation laws and project SHIELDS. Technical report, University of Patras.

Burton PW, Xu Y, Qin C, Tselentis GA, Sokos E. 2004a. A catalogue of seismicity in Greece and the adjacent areas for the twentieth century. *Tectonophysics* **390**: 117–127.

Burton PW, Xu Y, Qin C, Tselentis GA, Sokos E. 2004b. Extreme earthquake and earthquake perceptibility study in Greece and its surrounding area. *Natural Hazards* **32**: 277–312.

Burton PW, Xu Y, Tselentis GA, Sokos E, Aspinall W. 2003. Strong ground acceleration seismic hazard in Greece and neighbouring regions. *Soil Dynamics and Earthquake Engineering* **23**: 159–181.

Cabañas L, Benito B, Herráiz M. 1997. An approach to the measurement of the potential stuctural damage of earthquake ground motions. *Earthquake Engineering and Structural Dynamics* **26**: 79 – 92.

Calinski RB, Harabasz J. 1974. A dendrite method for cluster analysis. *Communications in Statistics* **3**: 1 – 27.

Campbell KW. 1997. Empirical near-source attenuation relationships for horizontal and vertical components of peak ground acceleration, peak ground velocity, and pseudo-absolute acceleration response spectra. *Seismological Research Letters* **68**(1): 154–179.

Campbell KW, Bozorgnia Y. 2006. Next Generation Attenuation (NGA) empirical ground motion models: can they be used in Europe? In: *Proceedings of the First European Conference on Earthquake Engineering and Seismology*. Geneva, Switzerland, p. Paper No. 458.

Campbell KW, Bozorgnia Y. 2007. Campbell and Bozorgnia NGA ground-motion relations for the geometric mean horizontal component of peak and spectral ground motion parameters. Peer report, Pacific Earthquake Engineering Research Center.

Chatzipetros A, Kokkalas S, Pavlides S, Koukouvelas I. 2005. Paleoseismic data and their implication for active deformation in Greece. *Journal of Geodynamics* **40**: 170–188.

Chiou BJ, Youngs RR. 2007. Chiou and Youngs PEER-NGA empirical ground motion model for the average horizonatal component of peak acceleration and pseudo-spectral acceleration for spectral periods of 0.01 to 10 seconds. Peer report, Pacific Earthquake Engineering Research Center.

Choi Y, Stewart JP. 2005. Nonlinear site amplification as function of 30m shear wave velocity. *Earthquake Spectra* **21**(1): 1 – 30.

Clarke PJ, Davies RR, England PC, Parsons BE, Billiris H, Paradissis D, Veis G, Denys PH, Cross PA, Ashkenazi V, Bingley R. 1997. Geodetic estimate of seismic hazard in the Gulf of Korinthos. *Geophysical Research Letters* **24**(11): 1303–1306.

Cocard M, Kahle HG, Peter Y, Geiger A, Veis G, Felekis S, Paradissis D, Billiris H. 1999. New constraints on the rapid crustal motion on the Aegean region: recent results inferred from GPS measurements (1993-1998) across the west Hellenic Arc. *Earth and Planetary Science Letters* **172**: 39–47.

Coley DA. 2005. *An introduction to genetic algorithms for scientists and engineers*. World Scientific, Singapore.

Coppersmith KJ, Youngs RR. 1986. Capturing uncertainty in probabilistic seismic hazard assessments within intraplate tectonic environments.

Cornell CA. 1968. Engineering seismic risk analysis. *Bulletin of the Seismological Society of America* **58**: 1583–1606.

Cotton F, Scherbaum F, Bommer JJ, Bungam H. 2006. Criteria for selecting and adjusting ground -motion models for specific target regions: Application to central Europe and rock sites. *Journal of Seismology* **10**: 137 – 156.

Cowie PA, Roberts GP. 2001. Constraining slip rates and spacings for active normal faults. *Journal of Structural Geology* **23**: 1901–1915.

Cramer CH, Petersen MD, Reichle MS. 1996. A monte carlo approach in estimating uncertainty for a seismic hazard assessment of Los Angeles, Ventura, and Orange Counties, California. *Bulletin of the Seismological Society of America* **86**(6): 1681–1691.

Danciu L, Tselentis GA. 2007. Engineering ground-motion parameters attenuation relationships for Greece. *Bulletin of the Seismological Society of America* **97**(1B): 162 – 183.

Davidon WC. 1991. Variable metric method for minimization. *SIAM Journal on Optimization* **1**: 1 – 17.

Douglas J. 2003. Earthquake ground motion estimation using strong-motion records: a review of equations for the estimation of peak ground acceleration and response spectral ordinates. *Earth-Science Reviews* **61**: 43–104.

Douglas J, Bertil D, Roulle A, Dominique P, Jousset P. 2006. A preliminary investigation of strong-motion data from the French Antilles. *Journal of Seismology* **10**(3): 271 – 199.

Drouet S, Scherbaum F, Cotton F, Souriau A. 2007. Selection and ranking of ground motion models for seismic hazard analysis in the Pyrenees. *Journal of Seismology* **11**(1): 87 – 100.

Du Q, Faber V, Gunzburger M. 1999. Centroidal Voronoi Tessellations. *SIAM Review* **41**(4): 637 – 676.

Ebel JE, Kafka AL. 1999. A Monte Carlo approach to seismic hazard analysis. *Bulletin of the Seismological Society of America* **89**(4): 854–866.

Engdahl ER, Hilst RVD, Buland R. 1998. Global teleseismic earthquake relocation with improved travel time and procedure for depth determination. *Bullettin of the Seismological Society of America* **88**: 722 – 743.

Erdik M, Biro YA, Onur T, Sesetyan K, Birgoren G. 1999. Assessment of earthquake hazard in Turkey and neighbouring regions. *Annali di Geofisica* **42**(6): 1125–1138.

FEMA. 2003. NEHRP recommended provisions for new buildings and other structures. Technical report, FEMA.

Feng Y, Hamerly G. 2006. PG-means: learning the number of clusters in data. In: *Advances in Neural Information Processing Systems 19*, Scholkopf B, Platt J, Hormann T (eds), MIT Press: Cambridge, MA, pp. 393 – 400.

Forgy E. 1965. Cluster analysis of multi-variate data: Efficiency vs. interpretability of classifications. *Biometrics* **21**: 768.

Frankel A. 1995. Mapping seismic hazard in the central and eastern United States. *Seismological Research Letters* **66**(4): 8–21.

Frankel A, Mueller C, Bernard T, Perkins D, Leyendecker E, Dickman N, Hansn S, Hopper M. 1996. National seismic hazard maps: Documentation June 1996. Technical Report 96-532, U.S. Geological Survey.

Frankel A, Petersen MD, Mueller CS, Haller KM, Wheeler RL, Leyendecker EV, Wesson R L Harmsen SC, Cramer CH, Perkins DM, Rukstales KM. 2005. Seismic hazard maps for the conterminous United States. Technical report, United States Geological Survey.

Frohlich C. 1989. The nature of deep-focus earthquakes. *Annual Reviews of Earth and Planetary Science* **17**: 227 – 54.

Frohlich C, Apperson KD. 1992. Earthquake focal mechanisms, moment tensors and consistency of seismic activity near plate boundaries. *Tectonics* **11**(2): 279 – 296.

Frohlich C, Davis SD. 1993. Teleseismic b values; or, much ado about 1.0. *Journal of Geophysical Research* **98**(B1): 631 – 644.

Ganas A, Drakatos G, Pavlidis SB, Stavrakakis GN, Ziazia M, Sokos E, Karastathis VK. 2005. The 2001 Mw=6.4 Skyros earthquake, conjugate strike-slip faulting and spatial variation in stress within the central Aegean Sea. *Journal of Geodynamics* **39**: 61–77.

Ganas A, Pavlides SB, Sboras S, Valkaniotis S, Papaioannou S, Alexandris GA, Plessa A, Papadopoulos GA. 2004. Active fault geometry and kinematics in Parnitha mountain, Attica, Greece. *Journal of Structural Geology* **26**: 2103–2118.

Gardner JK, Knopoff L. 1974. Is the sequence of earthquakes in Southern California, with aftershocks removed, Poissonian. *Bulletin of the Seismological Society of America* **64**: 1363 – 1367.

Gasperini P, Lolli B. 2006. Correlation between the parameters of the aftershock rate equation: Implications for the forecasting of future sequences. *Physics of the Earth and Planetart Interiors* **156**: 41 – 58.

Georgoudas IG, Sirakoulis GC, Scordilis EM, Andreadis I. 2007. A cellular automaton simulation tool for modelling seismicity in the region of Xanthi. *Environmental Modelling & Software* **22**: 1455 – 1464.

Gerstenberger MC, Wiemer S, Jones LM. 2004. Real-time forecasts of tomorrow's earthquakes in California: a new mapping tool. Technical Report 2004-1390, United States Geological Survey.

Gerstenberger MC, Wiemer S, Jones LM, Reasenberg PA. 2005. Real-time forecasts of tomorrow's earthquakes in california. *Nature* **435**: 328–331.

Giardini D, Wiemer S, Fäh D, Deichmann N. 2004. Seismic hazard assessment of Switzerland. Technical report, ETH Zurich.

Goldsworthy M, Jackson J, Haines J. 2002. The continuity of active fault systems in Greece. *Geophysical Journal International* **148**: 596–618.

Gringorten II. 1963. A plotting rule for extreme probability paper. *Journal of Geophysical Research* **68**: 813–814.

Grünthal G. 1998. European Macroseismic Scale 1998. Technical report, Conseil de l'Europe.

Grünthal G, Bosse C, Sellami S, Mayer-Rosa D, Giardini D. 1999. Compilation of the

GSHAP regional seismic hazard map for Europe, Africa and the Middle East,. *Annali di Geofisica* **42**: 1215 – 1223.

Grünthal G, Wahlström R. 2001. Sensitivity of parameters for probabilistic seismic hazard analysis using a logic tree approach. *Journal of Earthquake Engineering* **5**(3): 309 – 328.

Guo Z, Ogata Y. 1997. Statistical relations between the parameters of aftershocks in time, space, and magnitude. *Journal of Geophysical Research* **102**(B2): 2857 – 2873.

Gürer OF, Sangu E, Özburan M. 2005. Neotectonics of the SW Marmara region, NW Anatolia, Turkey. *Geological Magazine* **432**(2): 229–241.

Gutenberg B, Richter CF. 1944. Frequency of earthquakes in California. *Bulletin of the Seismological Society of America* **34**(4): 1985 – 1988.

Gutenberg B, Richter CF. 1956. Earthquake magnitude: intensity, energy and acceleration. *Bulletin of the Seismological Society of America* **46**: 104–145.

Hamerly G, Elkan C. 2003. Learning the k in k-means. In: *Advances in Neural Information Processing Systems 16*, Thrun S, Saul LK, Scholkopf B (eds), MIT Press: Cambridge, MA, pp. 281 – 288.

Hanks TC, Kanamori H. 1979. A moment-magnitude scale. *Journal of Geophysical Research* **84**: 2348 – 2350.

Hardy G, Merz K, Abrahamson N, Watson-Lamprey J. 2006. Program on technology innovation: Use of cumulative absolute velocity (CAV) in detemining effects of small magnitude earthquakes on seismic hazard analysis. Technical report, U.S. Department of Energy.

Hartigan JA. 1975. *Clustering algorithms*. John Wiley and Sons: New York, 1st edn.

Hartigan JA, Wong MA. 1979. Algorithm AS136: A k-means clustering algorithm. *Applied Statistics* **28**(1): 100–108.

Hatzfield D, Ziazia M, Kementzetzidou D, Hatzidimitriou P, Panagiotopoulos D, Makropoulos K, Papadimitriou P, Deschamps A. 1999. Microseismicity and focal mechanisms at the western termination of the North Anatolian Fault and their implications for continental tectonics. *Geophysical Journal International* **137**: 891–908.

Hatzidimitriou PM, Papadimitriou EE, Mountrakis DM, Papazachos BC. 1985. The seismic parameter b of the frequency-magnitude relation and its association with the geological zones in the area of Greece. *Tectonophysics* **120**: 141 –151.

Haupt RL, Haupt SE. 2004. *Practical genetic algorithms*. John Wiley & Sons, Hoboken, New Jersey.

Healy JH, Keilis-Borok VI, Lee WHK. 1997. Volume 6: Algorithms for earthquake statistics and prediction. Technical report, International Association of Seismology and Physics of the Earth's Interior.

Hollenstein C, Geiger A, Kahle HG, Veis G. 2006. CGPS time-series and trajectories of crustal motion along the West Hellenic Arc. *Geophysical Journal International* **164**: 182–191.

Idriss IM. 2007. Empirical model for estimating the average horizontal values of pseudo-spectral accelerations generated by crustal earthquakes (volume 1). Peer report, Pacific Earthquake Engineering Research Center.

Jackson D, Kagan YY. 2006. The 2004 Parkfield earthquake, the 1985 prediction, and characteristic earthquakes: Lessons for the future. *Bulletin of the Seismological Society of America* **96**(4B): S397 – S409.

Jackson J. 1999. Fault death: a perspective from actively deforming regions. *Journal of Structural Geology* **21**: 1003–1010.

Jain AK, Murty MN, Flynn PJ. 1999. Data clustering: A review. *ACM Computing Surveys* **31**(3): 264 – 322.

Jiménez A, Posadas AM. 2006. A moore's cellular automaton model to get probabilistic

seismic hazard maps for different magnitude releases: A case study for Greece. *Tectonophysics* **423**: 35 – 42.

Jiménez A, Posadas AM, Marfil JM. 2005. A probabilistic seismic hazard model based on cellular automata and information theory. *Nonlinear Processes in Geophysics* **12**: 1–16.

Jiménez MJ, Giardini D, Grunthal G, Erdik M, Garcia-Fernandez M, Lapajne J, Makropoulos K, Musson R, Papaioannou C, Rebez A, Riad S, Sellami S, Shapira A, Slejko D, van Eck T, El Sayed A. 2001. Unified seismic hazard modelling throughout the Mediterranean region. *Bollettino di Geofisica Teorica ed Applicata* **42**(1 - 2): 3 – 18.

Johnson CE, Koyanagi RY. 1988. A Monte-Carlo approach applied to the estimation of seismic hazard for the state of Hawaii. *Seismological Research Letters* **59**(1): 18.

Joyner WB, Boore DM. 1981. Peak horizontal acceleration and velocity from strong-motion records including records from the 1979 Imperial Valley, California, earthquake. *Bulletin of the Seismological Society of America* **71**(6): 2011–2038.

Kagan YY. 1993. Statistics of characteristic earthquakes. *Bulletin of the Seismological Society of America* **83**(1): 7–24.

Kalogeras IS, Burton PW. 1996. Shear-wave velocity models from Rayleigh-wave dispersion in the broader Aegean area. *Geophysical Journal International* **125**: 679–695.

Karacostas VG, Papadimitriou EE, Papazachos C. 2004. Properties of the 2003 Lefkada, Ionian Islands, Greece, earthquake seismic sequence and seismicity triggering. *Bulletin of the Seismological Society of America* **94**(5): 1976–1981.

Karakaisis GF, Hatzidimitriou PM, Scordilis EM, Panagiotopoulos DG. 1998. Seismicity of Western Macedonia, Greece. *Journal of Geodynamics* **26**(2-4): 297–307.

Kaufman L, Rousseeuw PJ. 1990. *Finding groups in data: An introduction to cluster analysis*. Wiley series in probability and mathematical statistics: Applied Probability and Statistics, John Wiley and Sons, Inc: United States of America, 1st edn.

Kayen RE, Mitchell JK. 1997. Arias intensity assessment of liquefaction test sites on the east side of San Francisco Bay affected by the Loma Prieta, California, earthquake of 17 October 1989. *Natural Hazards* **16**(2-3): 243–265.

Kijko A. 2004. Estimation of the maximum earthquake magnitude, Mmax. *Pure and Applied Geophysics* **161**: 1655–1681.

Kijko A, Sellevoll MA. 1989. Estimation of earthquake hazard parameters from incomplete data files. part i: Utilization of extreme and complete catalogues with different threshold magnitudes. *Bulletin of the Seismological Society of America* **79**(3): 645 – 654.

King GCP, Stein RS, Lin J. 1994. Static stress changes and the triggering of earthquakes. *Bulletin of the Seismological Society of America* **84**(3): 935–953.

Kiratzi A, Louvari E. 2003. Focal mechanisms of shallow earthquakes in the Aegean Sea and the surrounding lands determined by waveform modelling: a new database. *Journal of Geodynamics* **36**: 251–274.

Kiratzi AA. 2002. Stress tensor inversions along the westernmost North Anatolian Fault zone and its continuation into the North Aegean Sea. *Geophysical Journal International* **151**: 360–376.

Klügel JU. 2005. Problems in the application of the SSHAC probability method for assessing earthquake hazards at Swiss nuclear power plants. *Engineering Geology* **78**: 285 – 307.

Konstantinou KI, Kalogeras IS, Melis NS, Kourouzidis MC, Stavrakakis GN. 2006. The 8 January 2006 earthquake (Mw 6.7) offshore Kythira Island, Southern Greece: Seismological, strong-motion, and macroseismic observations of an intermediate-depth event. *Seismological Research Letters* **7**(5): 544 – 553.

Koravos GC, Main IG, Tsapanos TM, Musson RMW. 2003a. Maximum earthquake magnitudes in the Aegean area constrained by tectonic moment release rates. *Geophysical*

Journal International **152**: 94–112.

Koravos GC, Main IG, Tsapanos TM, Musson RMW. 2003b. Perceptible earthquakes in the broad Aegean area. *Tectonophysics* **371**: 175–186.

Kramer SL. 1996. *Geotechnical earthquake engineering*. Prentice-Hall International Series in Civil Engineering and Engineering Mechanics, Prentice Hall: New Jersey, 1st edn.

Kramer SL, Mitchell RA. 2006. Ground motion intensity measures for liquefaction hazard evaluation. *Earthquake Spectra* **22**(2): 413 – 438.

Krinitzsky EL. 2002. How to obtain earthquake ground motions for engineering design. *Engineering Geology* **65**: 1 – 6.

Krinitzsky EL, Chang FK. 1988. Intensity-related earthquake ground motions. *Bulletin of the Association of Engineering Geologists* **25**(4): 425 – 435.

Krishna K, Murty MN. 1999. Genetic K-means algorithm. *IEEE Transactions on Systems, Man and Cybernetics - Part B: Cybernetics* **29**(3): 433 – 439.

Krzanowski WJ, Lai YT. 1988. A criterion for determining the number of groups in a data set using sum-of-squares clustering. *Biometrics* **44**: 23 – 34.

Kuncheva LI, Vetrov DP. 2006. Evaluation of stability of k-means cluster ensembles with respect to random initialization. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28**(11): 1798 – 1808.

Le Pichon X, Angelier J. 1981. The Aegean Sea. *Philosophical Transactions of the Royal Society of London* **300**: 357–372.

Le Pichon X, Chamot-Rooke N, Rangin C, Segnör AMC. 2003. The North Anatolian fault in the Sea of Marmara. *Journal of Geophysical Research* **108**(B4): doi:10.1029/2002JB001 862.

Leeder MR, Portman C, Andrews JA, Collier REL, Finch E, Gawthorpe RL, McNeill LC, Perez-Arlucea M, Rowe P. 2005. Normal faulting and crustal deformation, Alkyonides Gulf and Perachora peninsula, eastern Gulf of Corinth rift, Greece. *Journal of the Geological Society of London* **162**: 549–561.

Likas A, Vlassis N, Verbeek JJ. 2003. The global k-means clustering algorithm. *Pattern Recognition* **36**: 451 – 461.

Lin PS, Lee CT. 2008. Ground-motion attenuation relationships for subduction-zone earthquakes in northeastern Taiwan. *Bulletin of the Seismological Society of America* **98**(1): 220 – 240.

Liu WB, Ma L. 2006. A recent application of the ETAS model and a proposed method for prediction of strong aftershocks. *Pure and Applied Geophysics* **163**: 2513 – 2528.

Lomnitz-Adler J. 1985. Asperity models and characteristic earthquakes. *Geophysical Journal of the Royal Astronomical Society* **83**: 435 – 450.

Lopez-Ruiz R, Vazquez-Prada M, Gomez JB, Pacheco A. 2004. A model of characteristic earthquakes and its implications for regional seismicity. *Terra Nova* **16**(3): 116 – 120.

Louvari E, Kiratzi A, Papazachos B, Hatzidimitriou P. 2001. Fault-plane solutions determined by waveform modelling confirm tectonic collision in the eastern Adriatic. *Pure and Applied Geophysics* **158**: 1613–1637.

Louvari E, Kiratzi AA, Papazachos BC. 1999. The Cephalonia Transform Fault and its extension to western Lefkada Island (Greece). *Tectonophysics* **308**: 223–236.

Lu Y, Lu S, Foyouhi F, Deng Y, Brown SJ. 2004. Incremental genetic K-means algorithm and its application to gene expression data analysis. *BMC Bioinformatics* **5**: doi:10.1186/1471–2105–5–172.

Luhr J. 2006. Volcanoes of the Mediterranean and western Asia.

Main IG, Burton PW. 1984. Information theory and the earthquake frequency-magnitude distribution. *Bulletin of the Seismological Society of America* **74**(4): 1409–1426.

Main IG, Burton PW. 1989. Seismotectonics and the earthquake frequency-magnitude

distribution in the Aegean area. *Geophysical Journal* **98**: 575–586.

Makropoulos KC, Burton PW. 1981. A catalogue of seismicity in Greece and adjacent areas. *Geophysical Journal of the Royal Astronomical Society* **65**: 741 – 762.

Makropoulos KC, Burton PW. 1983. Seismic risk of circum-Pacific earthquakes i: Strain-energy release. *Pure and Applied Geophysics* **121**(2): 247–267.

Makropoulos KC, Burton PW. 1984. Greek tectonics and seismicity. *Tectonophysics* **106**: 275–304.

Makropoulos KC, Burton PW. 1985a. Seismic hazard in Greece I: Magnitude recurrance. *Tectonophysics* **117**: 205–257.

Makropoulos KC, Burton PW. 1985b. Seismic hazard in Greece II: Ground acceleration. *Tectonophysics* **117**: 259–294.

Margaris B, Papazachos C, Papaioannou C, Theodulidis N, Kalogeras I, Skarlatoudis A. 2001. Ground motion attenuation relations for shallow earthquakes in Greece. In: *12th European Conference on Earthquake Engineering*.

Margaris BN, Hatzidimitriou PM. 2002. Source spectral scaling and stress release estimates using strong-motion records in Greece. *Bulletin of the Seismological Society of America* **92**(3): 1040 – 1059.

Margaris BN, Papazachos CB. 1999. Moment-magnitude relations based on strong-motion records in Greece. *Bullettin of the Seismological Society of America* **89**: 442 – 455.

Marsaglia G, Tsang WW. 2000. The ziggarut method for generating random variables. *Journal of Statistical Software* **5**(8): 1 – 7.

Marsaglia G, Zaman A. 1991. A new class of random number generators. *Annals of Applied Probability* **3**: 462–480.

Matthews MV, Ellsworth WL, Reasenberg PA. 2002. A Brownian model for recurrent earthquakes. *Bulletin of the Seismological Society of America* **92**(6): 2233–2250.

McClusky S, Balassanian S, Barka A, Demir C, Ergintav S, Georgiev I, Gurkan O, Hamburger M, K H, Kahle H, Kastens K, Kekelidze G, King R, Kotzev V, Lenk O, Mahmoud S, Mishin A, Nadariya M, Ouzounis A, Paradissis D, Peter Y, Prilepin M, Reilenger R, Sanli I, Seegher H, Tealeb A, Toksöz MN, Veis G. 2000. Global positioning system constraints on plate kinematics and dynamics in the eastern Mediterranean and Caucasus. *Journal of Geophysical Research* **105**(B3): 5695–5720.

McGuire RK. 1976. FORTRAN computer program for seismic risk analysis. Technical report, USGS.

McGuire RK. 1995. Probabilistic seismic hazard analysis and design earthquakes: Closing the loop. *Bulletin of the Seismological Society of America* **85**(5): 1275–1284.

McKenzie D. 1978. Active tectonics of the Alpine-Himalayan belt: the Aegean Sea and surrounding regions. *Geophysical Journal of the Royal Astronomical Society* **55**: 217–254.

McNeill LC, Mille A, Minsull TA, Bull JM, Kenyon NH. 2004. Extension of the North Anatolian Fault into the North Aegean Trough: Evidence for transtension, strain partitioning, and analogues for Sea of Marmara basin models. *Tectonics* **23**(TC2016): doi:10.1029/2002TC001 490.

Meijer PT, Wortel MJR. 1997. Present-day dynamics of the Aegean region: A model analysis of the horizontal pattern of stress and deformation. *Tectonics* **16**(6): 879–895.

Milligan GW, Cooper MC. 1985. An examination of procedures for determining the number of clusters in a data set. *Psychometrika* **50**(2): 159 – 179.

Mogi K. 1962. Magnitude-frequency relationship for elastic shocks accompanying fractures of various materials and some related problems in earthquakes. *Bullettin of the Earthquake Research Institute., Tokyo University* **40**: 831 – 883.

Montaldo V, Faccioli E, Zonno G, Akinci A, Malagnini L. 2005. Treatment of ground-motion predictive relationships for the reference seismic hazard map of Italy. *Journal of*

Seismology **9**: 295 – 316.

Moretti I, Sakellariou D, Lykousis V, Micarelli L. 2003. The Gulf of Corinth: an active half graben? *Journal of Geodynamics* **36**: 323–340.

Muço B. 1994. Focal mechanism solutions for Albanian earthquakes for the years 1964-1988. *Tectonophysics* **231**: 311–323.

Musson RMW. 1999a. Determination of design earthquakes in seismic hazard analysis through Monte Carlo simulation. *Journal of Earthquake Engineering* **3**(4): 463–474.

Musson RMW. 1999b. Probabilistic seismic hazard maps for the North Balkan region. *Annali di Geofisica* **42**(6): 1109–1124.

Musson RMW. 2000. The use of Monte Carlo simulations for seismic hazard assessment in the U.K. *Annali di Geofisica* **43**(1): 1–9.

Musson RMW. 2004. Objective validation of source models. In: *13th World Conference on Earthquake Engineering*. Vancouver, Canada, p. Paper No. 2492.

Musson RMW. 2005. Against fractiles. *Earthquake Spectra* **21**(3): 887 – 891.

Musson RMW. 2009. Ground motion and probabilistic hazard. *Bulletin of Earthquake Engineering* **In Press**.

Musson RMW, Henni PHO. 2001. Methodological considerations of probabilistic seismic hazard mapping. *Soil Dynamics and Earthquake Engineering* **21**: 385–403.

Musson RMW, Sargeant SL. 2007. Eurocode 8 seismic hazard zoning maps for the U.K. Technical report, British Geological Survey.

Musson RMW, Tsapanos T, Naka CT. 2002. A power-law function for earthquake interarrival time and magnitude. *Bulletin of the Seismological Society of America* **92**(5): 1783–1794.

Nabney IT. 2002. *Netlab: Algorithms for pattern recognition*. Advances in Pattern Recognition, Springer: London, 1st edn.

Nyst M, Thatcher W. 2004. New constraints on the active tectonic deformation of the Aegean. *Journal of Geophysical Research* **109**(B11406): doi:10.1029/2003JB002 830.

Ogata Y. 1983. Estimation of the parameters in the modified Omori formula for aftershock frequencies by the maximum likelihood procedure. *Journal of the Physics of the Earth* **31**(2): 115 – 124.

Ogata Y. 1988. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association* **83**(401): 9 – 27.

Ogata Y. 1993. Fast likelihood computation of epidemic type aftershock-sequence model. *Geophysical Research Letters* **20**(19): 2143 – 2143.

O'Hara TF, Jacobson JP. 1991. Standardization of the Cumulative Absolute Velocity. Technical report, Electric Power Research Institute.

Okay AI, Tüysüz O, Kaya c. 2004. From transpression to transtension: changes in morphology and structure around a bend in the North Anatolian fault in the Marmara region. *Tectonophysics* **391**: 259–282.

Ordaz M, Reyes C. 1999. Earthquake hazard in Mexico City: Observations versus computations. *Bulletin of the Seismological Society of America* **89**(5): 1379 – 1383.

Page R. 1968. Aftershocks and microaftershocks. *Bulletin of the Seismological Society of America* **58**: 323 – 335.

Papadimitriou E. 2002. Mode of strong earthquake recurrence in the Central Ionian Islands (Greece): Possible triggering due to coulomb stress changes generated by the occurrence of previous strong shocks. *Bulletin of the Seismological Society of America* **92**(8): 3293–3308.

Papadimitriou P, Kaviris G, Makropoulos K. 2006. The Mw = 6.3 2003 Lefkada earthquake (Greece) and induce stress transfer changes. *Tectonophysics* **423**: 73–82.

Papadopoulos GA, Drakatos G, Papanastassiou D, Kalogeras I, Stavrakakis G. 2000. Pre-

liminary results about the catastrophic earthquake of 7 September 1999 in Athens, Greece. *Seismological Research Letters* **71**(3): 318 – 329.

Papadopoulos GA, Ganas A, Plessa A. 2002. The Skyros earthquake (Mw 6.5) of 26 July 2001 and precursory seismicity patterns in the North Aegean Sea. *Bulletin of the Seismological Society of America* **92**(3): 1141 – 1145.

Papadopoulos GA, Karastathis VK, Ganas A, Pavlidis S, Fokaefs A, Orfanogiannaki K. 2003. The Lefkada, Ionian Sea (Greece), shock (Mw 6.2) of 14 August 2003: Evidence for the characteristic earthquake from seismicity and ground failures. *Earth, Planets and Space* **55**(11): 713 –718.

Papadopoulos GA, Skafida H, Vassiliou IT. 1993. Nonlinearity of the magnitude-frequency relation in the Hellenic Arc-Trench system and the characteristic earthquake model. *Journal of Geophysical Research* **98**(B10): 17,737 – 17,744.

Papaioannou CA, Papazachos BC. 2000. Time-independent and time-dependent seismic hazard in Greece based on seismogenic sources. *Bulletin of the Seismological Society of America* **90**(1): 22–33.

Papazachos BC. 1990. Seismicity of the aegean and surrounding area. *Tectonophysics* **178**: 287–308.

Papazachos BC, Kiratzi AA, Hatzidimitriou PM, Rocca AC. 1984. Seismic faults in the Aegean area. *Tectonophysics* **106**: 71–85.

Papazachos BC, Kiratzi AA, Papadimitriou E. 1991. Regional focal mechanisms for earthquakes in the aegean area. *Pure and Applied Geophysics* **136**(4): 405–432.

Papazachos BC, Papaioannou CA, Papazachos CB, Savvidis AS. 1997. *Atlas of isoseismal maps for strong shallow earthquakes in Greece and surrounding area (426BC - 1995)*. Geophysical Laboratory Publication No. 4, P. Ziti and Co.: Thessaloniki.

Papazachos BC, Papaioannou CA, Papazachos CB, Savvidis AS. 1999. Rupture zones in the Aegean region. *Tectonophysics* **308**: 205–221.

Papazachos BC, Papazachou C. 1997. *The earthquakes of Greece*. Technical books Editions, P. Ziti and co.: Thessaloniki, Greece, 1 edn.

Papazachos CB. 1999a. An alternative method for a reliable estimation of seismicity with an application in Greece and the surrounding area. *Bulletin of the Seismological Society of America* **89**(1): 111 – 119.

Papazachos CB. 1999b. Seismological and GPS evidence for the Aegean-Anatolia interaction. *Geophysical Research Letters* **26**(17): 2653–2656.

Papazachos CB, Kiratzi AA. 1996. A detailed study of active deformation in the Aegean and surrounding area. *Tectonophysics* **253**: 129–153.

Papazachos CB, Kiratzi AA, Papazachos BC. 1992. Rates of active crustal deformation in the Aegean and the surrounding area. *Journal of Geodynamics* **16**(3): 147–179.

Papoulia J, Makris J, Drakopoulou V. 2006. Local seismic array observations at north Evoikos, central Greece, delineate crustal deformation between the North Aegean Trough and Corinthiakos Rift. *Tectonophysics* **423**: 97–106.

Parsons T. 2004. Recalculated probability of M ¿ 7 earthquakes beneath the Sea of Marmara, Turkey. *Journal of Geophysical Research* **109**(B05304): doi:10.1029/2003JB002 667.

Parsons T, Toda S, Stein RS, Barka A, Dietrich JH. 2000. Heightened odds of large earthquakes near Istanbul: An interaction-based probability calculation. *Science* **288**: 661–665.

Pavlides S, Papadopoulos G, Ganas A. 2002. The fault that caused the Athens September 1999 Ms = 5.9 earthquake: Field observations. *Natural Hazards* **27**: 61–84.

Peña JM, Lozano JA, Larranaga P. 1999. An empirical comparison of four initialization methods for the K-means algorithm. *Pattern Recognition Letters* **20**: 1027 – 1040.

Petersen MD, Frankel AD, Harmsen SC, Mueller CS, Haller KM, Wheeler RL, Wesson

RL, Zeng Y, Boyd OS, Perkins DM, Luco N, Field EH, Wills CJ, Rukstales KS. 2008. Documentation for the 2008 update of the United States national seismic hazard maps. *U.S. Geological Survey Open-File Report* **1268**: 1–61.

Pinar A, Kuge K, Honkura Y. 2003. Moment tensor inversion of recent small to moderate sized earthquakes: implications for seismic hazard and active tectonics beneath the Sea of Marmara. *Geophysical Journal International* **153**: 133–145.

Reasenberg P. 1985. Second-order moment of Central California seismicity, 1969 - 1982. *Journal of Geophysical Research* **90**(B7): 5479–5495.

Reasenberg P, Jones LM. 1989. Earthquake hazard after a mainshock in California. *Science* **243**: 1173–1176.

Reilinger R, Ergintav S, Burgmann R, McClusky S, Lenk O, Barka A, Gurkan O, Hearn L, Feigl KL, Cakmak R, Aktug B, Ozener H, Toksöz MN. 2000. Coseismic and postseismic fault slip for the 17 August 1999, M = 7.5, Izmit, Turkey earthquake. *Science* **289**(5484): 1.

Reilinger R, McClusky S, Vernant P, Lawrence S, Ergintav S, Cakmak R, Ozener H, Kadirov F, Guliev I, Stepanyan R, Nadariya M, Hahubia G, Mahmoud S, Sakr K, ArRajehi A, Paradissis D, Al-Aydrus A, Prilepin M, Guseva T, Evren E, Dmittrotsa A, Filikov SV, Gomez F, Al-Ghazzi R, Karam G. 2006. GPS constraints on continental deformation in the Africa-Arabia-Eurasia continental collision zone and implications for the dynamics of plate interactions. *Journal of Geophysical Research* **111**(B05411): doi:10.1029/2005JB004 051.

Reiter L. 1990. *Earthquake hazard analysis: Issues and insights*. Columbia University Press, New York.

Resor PG, Pollard DD, Wright TJ, Beroza GC. 2005. Integrating high-precision aftershock locations and geodetic observations to model coseismic deformation associated with the 1995 Kozani-Grevena earthquake, Greece. *Journal of Geophysical Research* **110**(B9): doi:10.1029/2004JB003 263.

Richter CF. 1958. *Elementary seismology*. W. H. Freeman, San Francisco, CA.

Rotondi R, Varini E. 2006. Bayesian analysis of marked stress release models for time-dependent hazard assessment in the western Gulf of Corinth. *Tectonophysics* **423**: 107–113.

Rotondi R, Varini E. 2007. Bayesian inference of stress release models applied to some Italian seismogenic zones. *Geophysical Journal International* **169**(1): 301 – 314.

Scherbaum F, Bommer JJ, Bungam H, Cotton F, Abrahamson NA. 2005. Composite ground-motion models and logic trees: Methodology, sensitivities, and uncertainties. *Bulletin of the Seismological Society of America* **95**(5): 1575 – 1593.

Scherbaum F, Cotton F, Smit P. 2004a. On the use of response spectral-reference data for the selection and ranking of ground-motion models for seismic hazard analysis in regions of moderate seismicity: The case of rock motion. *Bulletin of the Seismological Society of America* **94**(6): 2164 –2185.

Scherbaum F, Schmedes J, Cotton F. 2004b. On the conversion of source-to-site distance measures for extended earthquake source models. *Bulletin of the Seismological Society of America* **94**(3): 1053 – 1069.

Schurch P, Becker E. 2005. Studies on 'precarious rocks' in the epicentral area of the AD 1356 Basle earthquake, Switzerland. *Geophysical Journal International* **163**(2): 689 – 697.

Schwartz D, Coppersmith KJ. 1984. Fault behaviour and characteristic earthquakes: Examples from the Wasatch and San Andreas Fault Zones. *Journal of Geophysical Research* **89**(B7): 5681 – 5698.

Segnor AMC, Tüysüz O, Imran C, Sakinç M, Eyidoğan H, Görür N, Le Pichon X, Rangin C. 2005. The North Anatolian Fault: A new look. *Annual Reviews of Earth and Planetary*

Science **33**: 37–112.

Shapira A. 1983. Potential earthquake risk estimations by application of a simulation process. *Tectonophysics* **95**: 75–89.

Shcherbakov R, Turcotte DL, Rundle JB. 2005. Aftershock statistics. *Pure and Applied Geophysics* **162**: 1051 – 1076.

Shebalin NV, Karnik V, Hadzvieski D. 1974. Balkan region - a catalogue of earthquakes. Technical report, UNESCO Project Office.

Sheng W, Liu X. 2006. A genetic k-medoids clustering algorithm. *Journal of Heuristics* **12**: 447 – 466.

Shi Y, Bolt BA. 1982. The standard error of the magntiude-frequency b value. *Bulletin of the Seismological Society of America* **72**(5): 1677–1687.

Sinadinovski C, Edwards M, Corby N, Milne M, Dale K, Dhu T, Jones A, McPherson A, Jones T, Gray D, Robinson D, White J. 2005. Earthquake risk. In: *Natural hazard risk in Perth, WA*, Jones T, Middelmann M, Corby N (eds), Geoscience Australia: Canberra, Australia.

Skarlatoudis AA, Papazachos CB, Margaris BN, Theodolidis N, Papaioannou C, Kalogeras I, Scordilis EM, Karakostas V. 2003. Empirical peak ground motion predictive relationships for shallow earthquakes in Greece. *Bulletin of the Seismological Society of America* **93**(6): 2591–2603.

Slejko D, Camassi R, Cecić I, Herak D, Herak M, Kociu S, Kouskouna V, Lapajne J, Makropoulos K, Meletti C, Muço B, Papaioannou C, Peruzza L, Rebez A, Scandone P, Sulstarova E, Voulgaris N, Živčić M, Zupančić P. 1999. Seismic hazard assessment for Adria. *Annali di Geofisica* **42**(6): 1085–1107.

Smith SW. 1976. Determination of maximum earthquake magnitude. *Geophysical Research Letters* **3**(6): 351–354.

Smith WD. 2003. Earthquake hazard and risk assessment in New Zealand by Monte Carlo methods. *Seismological Research Letters* **74**(3): 298–304.

Spudich P, Joyner WB, Lindh AG, Boore DM, Margaris BM, Fletcher JB. 1999. SEA99: A revised ground motion prediction relation for use in extensional tectonic regimes. *Bulletin of the Seismological Society of America* **89**(5): 1156 – 1170.

Stafford PJ, Berrill JB, Perringa JR. 2008a. New predictive equations for Arias intensity from crustal earthquakes in New Zealand. *Journal of Seismology* : in press.

Stafford PJ, Strasser FO, Bommer JJ. 2008b. An evaluation of the applicability of the NGA models to ground-motion prediction in the Euro-Mediterranean region. *Bulletin of Earthquake Engineering* **6**(2): 149 – 177.

Stein RS, Barka AA, Dietrich JH. 1997. Progressive failure on the North Anatolian fault since 1939 by earthquake stress triggering. *Geophysical Journal International* **128**: 594–604.

Stein S, Newman A. 2004. Characteristic and uncharacteristic earthquakes as possible artifacts: application to the New Madrid and Wabash seismic zones. *Seismological Research Letters* **75**: 2173 – 2187.

Stepp JC. 1971. An investigation of earthquake risk in the Puget Sound area by use of the type I distribution of largest extremes. PhD thesis, Pennsylvania State Univeristy.

Stewart I. 1996. Holocene uplift and paleoseismicity on the Eliki Fault, Western Gulf of Corinth, Greece. *Annali di Geofisica* **39**(3): 575–588.

Stirling MW, Anooshehpoor A. 2006. Constraints on probabilistic seismic hazard models from unstable landform features in New Zealand. *Bulletin of the Seismological Society of America* **96**(2): 404–414.

Stirling MW, McVerry GH, Berryman KR. 2002. A new seismic hazard model for New Zealand. *Bulletin of the Seismological Society of America* **92**(5): 1878 – 1903.

Stiros SC. 1998. Historical seismicity, paeloseismicity and seismic risk in Western Macedonia, Northern Greece. *Journal of Geodynamics* **26**(2-4): 271–287.

Strasser FO, Bommer JJ, Abrahamson NA. 2008. Truncation of the distribution of ground -motion residuals. *Journal of Seismology* **12**: 79–105.

Sulstarova E, Peçi V, Shuteriqi P. 2000. Vlori-Elbasani-Dibra (Albania) transversal fault zone and its seismic activity. *Journal of Seismology* **4**: 117–131.

TERA. 1980. Seismic hazard analysis: a methodology for the eastern United States. Technical report, TERA Corporation.

Theodulidis NP, Papazachos BC. 1992. Dependence of strong ground motion on magnitude-distance, site geology and macroseismic intensity for shallow earthquakes in Greece: I, peak horizontal acceleration, velocity and displacment. *Soil Dynamics and Earthquake Engineering* **11**: 387–402.

Tibshirani R, Walther G, Hastie T. 2001. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society (B)* **63**(2): 411 – 423.

Tranos MD, Papadimitrou EE, Kilias AA. 2003. Thessaloniki-Gerakarou Fault Zone (TGFZ): the western extension of the 1978 Thessaloniki earthquake fault (Northern Greece) and seismic hazard assessment. *Journal of Structural Geology* **25**: 2109–2123.

Travasarou T, Bray JD, Abrahamson NA. 2003. Empirical attenuation relationship for Arias Intensity. *Earthquake Engineering and Structural Dynamics* **32**: 1133 – 1155.

Trifunac MC, Brady AG. 1975. On the correlation of seismic intensity with peaks of recorded strong ground motion. *Bulletin of the Seismological Society of America* **65**: 139 –162.

Tselentis GA, Danciu L. 2008. Empirical relationships between Modified Mercalli Intensity and engineering ground-motion parameters in Greece. *Bulletin of the Seismological Society of America* **98**: 1863 – 1875.

Turcotte DL. 1997. *Fractals and chaos in geology and geophysics*. Cambridge University Press: Cambridge, United Kingdom.

Utsu T. 1961. A statistical study on the occurrence of aftershocks. *Geophysical Magazine* **30**: 521 – 605.

Utsu T. 1965. A method for determining the value of b in a formula log n = a - bM showing the magnitude-frequency relations for earthquakes. *Geophysical Bullettin of Hokkaido Univeristy* **13**: 99 – 103.

Utsu T. 1999. Representation and analysis of the earthquake size distribution: A historical review and some new approaches. *Pure and Applied Geophysics* **155**: 509–535.

Vamvakaris DA, Papazachos CB, Karagianni EE, Scordilis EM, Hatzidimitriou PM. 2006. Small-scale spatial variation of the stress field in the back-arc Aegean area: Results from the seismotectonic study of the broader area of Mygdonia basin (N. Greece). *Tectonophysics* **417**: 249–267.

Vannucci G, Pondrelli S, Argnani A, Morelli A, Gasperini P, Boschi E. 2004. An atlas of Mediterrean seismicity. *Annals of Geophysics* **Supplement to 47**(1): 247–306.

Wald DJ, Allen TI. 2007. Topographic slope as a proxy for seismic site conditions and amplification. *Bulletin of the Seismological Society of America* **97**(5): 1379 – 1395.

Wald DJ, Quitoriano V, Heaton TJ, Kanamori H. 1999. Relationships between peak ground acceleration, peak ground velocity and Modified Mercalli Intensity in California. *Earthquake Spectra* **15**: 557 – 564.

Wald DJ, Worden BC, Quitoriano V, Pankow KL. 2005. ShakeMap(R) Manual: Technical manual, users guide and software guide. Technical report, United States Geological Survey (USGS).

Weatherill GA, Burton PW. 2009. Delineation of shallow seismic sources using K-means cluster analysis, with specific application to the Aegean region. *Geophysical Journal In-*

ternational **176**(2): 565 – 588.

Welling M, Kurihara K. 2006. Bayesian K-means as a "maximization-expectation" algorithm. In: *Proceedings of the 2006 SIAM Conference on Data Mining*. Society for Industrial and Applied Mathematics, Philadelphia, PA: Bethesda, Maryland, pp. 474 – 478,.

Wells DL, Coppersmith KJ. 1994. New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement. *Bulletin of the Seismological Society of America* **84**(4): 974–1002.

Wesnousky SG. 1994. The Gutenberg-Richter or characteristic earthquake distribution, which is it? *Bulletin of the Seismological Society of America* **84**(6): 1940–1959.

Wiemer S. 2001. A software package to analyze seismicity: ZMAP. *Seismological Research Letters* **72**(2): 374–383.

Wiemer S, Giardini D, Fäh D, Deichmann N, Sellami S. 2009. Probabilistic seismic hazard assessment of Switzerland: best estimates and uncertainties. *Journal of Seismology* **in press**.

Wiemer S, Wyss M. 2000. Minimum magnitude of completeness in earthquake catalogs: Examples from Alaska, the Western United States, and Japan. *Bulletin of the Seismological Society of America* **90**(4): 859–869.

Wills CJ, Clahan KB. 2006. Developing a map of geologically defined site-condition categories for California. *Bulletin of the Seismological Society of America* **96**(4A): 1483 – 1501.

Woo G. 1996. Kernel estimation methods for seismic hazard area source modeling. *Bulletin of the Seismological Society of America* **86**(2): 353–362.

Wössner J, Wiemer S. 2005. Assessing the quality of earthquake catalogues: Estimating the magnitude of completeness and its uncertainty. *Bulletin of the Seismological Society of America* **95**(2): 684–698.

Wright T, Fielding E, Parsons B. 2001. Triggered slip: Observations of the 17 August 1999 Izmit (Turkey) earthquake using radar interferometry. *Geophysical Research Letters* **28**(6): 1079 – 1082.

Xie XL, Beni G. 1991. A validity measure for fuzzy clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **13**(8): 841–847.

Xu Y, Burton PW. 2001. Rescaled range analysis of the frequency of occurrance of moderate-strong earthquakes in the Mediterranean area. In: *Emergent Nature: Patterns, Growth and Scaling in the Sciences.*, Novak M (ed), World Scientific Publishing Co.: London, pp. 305–314.

Xu Y, Burton PW. 2006. Time varying seismicity in Greece: Hurst's analysis and Monte Carlo simulation applied to a new earthquake catalogue for Greece. *Tectonophysics* **423**(1-4): 125 – 136.

Youngs RR, Abrahamson N, Makdisi FI, Sadigh K. 1995. Magnitude-dependent variance of peak ground acceleration. *Bulletin of the Seismological Society of America* **85**(4): 1161 – 1176.

Youngs RR, Chiou SJ, Silva WJ, Humphrey JR. 1997. Strong ground motion attenuation relationships for subduction zone earthquakes. *Seismological Research Letters* **68**(1): 58 – 73.

Youngs RR, Coppersmith KJ. 1985. Implications of fault slip rates and earthquake recurrance models to probabilistic seismic hazard estimates. *Bulletin of the Seismological Society of America* **75**(4): 939–964.

Zamani A, Hashemi N. 2004. Computer-based self-organized tectonic zoning: a tentative pattern recognition for Iran. *Computers and Geosciences* **30**: 705–718.

Zanchi A, Angelier J. 1993. Seismotectonics of western Anatolia: regional stress orientation from geophysical and geological data. *Tectonophysics* **222**: 259–274.

Zhang B, Hsu M, Dayal U. 1999. K-Harmonic means: A data clustering algorithm. Tech-

nical report, Hewlitt Packard Research Laboratory.

Appendix A

Development of a New Seismic Source Model

The novel source model presented in Chapter 3 is derived from an interpretation of a variety of seismotectonic information available at the time of its development (October/November 2006). As with other zone models that are delineated according to expert judgement, explanation for the locations of particular boundaries is required. As discussed in Chapter 3, the justification for the location of a particular zone is not always made clear, even in widely used zone models. For the purposes of transparency, the following appendix is a report that details the decisions made in the delineation of the source zones in the WT2006 model. References are included in the bibliography of the main body of this thesis.

A.1 Western Greece, the Balkans and the Adriatic Coast

This region extends along the Adriatic coast from southern Croatia to Corfu. It includes Albania, western Macedonia and northwest Greece. As with the Hellenic Arc, this is a region of tectonic compression; here manifest not as a subduction zone but as an orogenic belt. Driven by the continental scale Africa-Eurasia collision, the seismicity of this region displays predominantly thrust characteristics (Anderson and Jackson, 1987). In the Adriatic Balkans and western Greece, high seismicity can be readily attributed this compressional faulting environment.

A simple tectonic model of the Adriatic is seen in Figure A.1. This formed the basis for the Adriatic source zone model used in the GSHAP project (Slejko *et al.*, 1999), Analysis of focal mechanisms within this region show clear thrust faulting with strikes of 310° N $-$ 340° N and dips of 10° $-$ 35° (Baker *et al.*, 1997). Several large damaging earthquakes have been recorded in this area, including that of 13 June 1563 (M_W 7), 6 April 1667

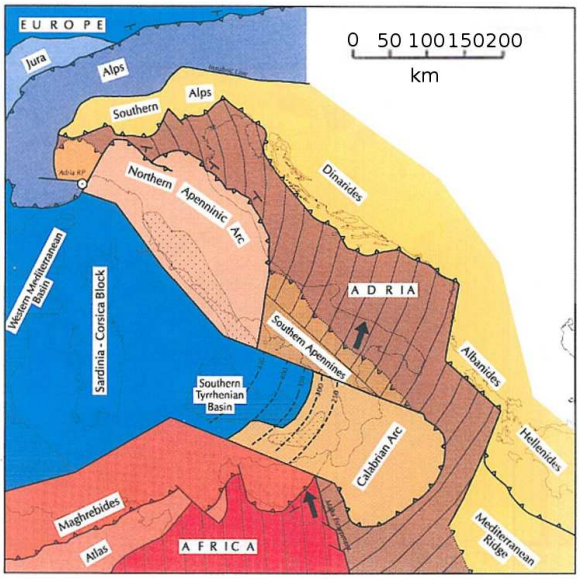

Figure A.1: Slejko *et al.* (1999) tectonic model of the Adriatic region.

(M_W 7.2) and 15 April 1979 (M_W 6.9).

A large cluster of seismicity occurs within a small area around the coastal section of the Montenegro-Albanian border. Seismotectonic maps show a series of thrust faults, striking parallel to the line of the Dalmatian coast. The trace of these faults into Albania is less clear. Reilinger *et al.* (2006) [REIL2006] suggest that this region represents a triple junction between an Adriatic block, the main Eurasian block and a southwest Balkan block. This would manifest as a series of thrust faults extending northwest up the Adriatic coast and further south (through western Albania). This can be seen in Figure A.2. The rationale behind the delineation of the southern Eurasian boundary is to define of a single coherent Black Sea tectonic block. This edge extends east across the Black Sea before terminating in the Caucasus region of eastern Turkey. This boundary is poorly defined by geodetic data and displays little seismic activity.

In the REIL2006 block model the slip rates for the block margins described are low. The NW striking margin originating in Montenegro is almost entirely compressional, slipping at a rate of 3.4 - 4.2 mm/yr. The sinistral strike-slip component of motion here is practically negligible. This is greater than the velocity calculated by Papazachos *et al.* (1992) of 2.25 mm/yr, which is derived from annual seismic moment rate in the region. For the NE striking margin, REIL2006 propose an oblique mechanism with approximately 3.5 mm/yr extension and 1.3 mm/yr dextral strike-slip. This change to an extensional regime is most likely a result of the block rotation required to match the slip rates further east in the block.

Whilst the tectonics of the Dalmatian coast appears to be adequately described by a model of orogenic collision between the Adriatic lithosphere and Eurasian lithosphere,

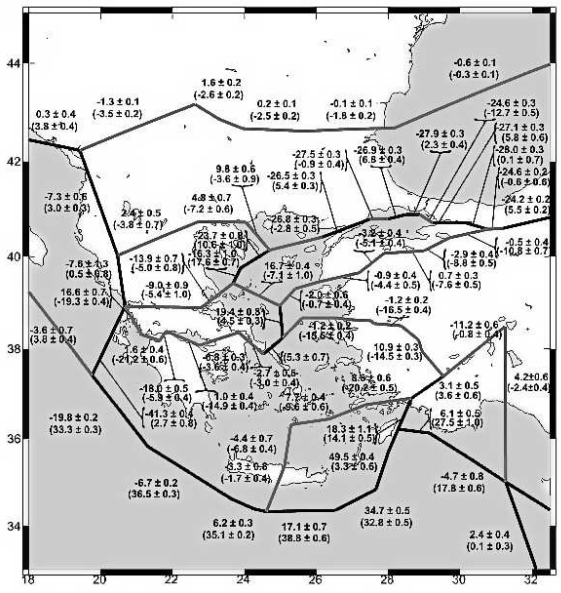

Figure A.2: Aegean tectonic block model of REIL2006. Slip rates: strike-slip (no brackets sinistral positive) and normal (brackets - compression positive), grey faults are those poorly constrained by geodetic data.

the situation in Albania is more complex. Faulting in this region displays a consistent pattern of low angle (with a dip of 25 - 30°) thrust movement along the western Albanian coastline. Focal mechanisms of earthquakes in this region indicate nodal planes with strikes in the NNW direction. This is consistent with the compression arising from the Adriatic-Eurasian collision, though the transition from NE to NNE faulting is indicative of a change in the fault strike. Focal mechanisms of earthquakes in eastern Albania, however, clearly demonstrate extensional faulting (Anderson and Jackson, 1987; Baker *et al.*, 1997; Louvari *et al.*, 2001).

Anderson and Jackson (1987) give little insight into the cause of the largest normal faulting events, suggesting only that they represent the transition from the orogenic regime in the West, to the extensional regime that characterises most of central Greece. Baker *et al.* (1997) extend this hypothesis, noting the difference in strike between the ENE striking faults in eastern Albania and the clearer N-S extension observed in central Greece. This is consistent with the rotation of a Central Greece and South Aegean two-microplate system, inferred from GPS vectors by Nyst and Thatcher (2004). The block model suggested by Reilinger *et al.* (2006) also supports this hypothesis. This is visible in the dextral strike-slip motion along a fault striking WNW across Albania, and an extensional fault with a sinistral strike-slip component in Northern Greece. Louvari *et al.* (2001) produce alternative hypotheses, having postulated over a number of different theories as to why E-W tension may be observed. One possibility is that orogenic compression gives rise to a dense lithospheric downbulge, which forces the lower density crustal root beneath the mountain range (Bott, 1993). The low-density crustal root is then assumed to be responsible for extension on the lee flank of the mountain range. A full treatment of the

geodynamic reasons for extensional faulting in a compressional environment are beyond the scope of this discussion.

Crustal deformation rates within the Albanian region reflect the contrasting regimes described previously. Using the PZ1990 model, Papazachos *et al.* (1992) estimated deformation rates for four zones in the Albania-Macedonia-Greece region. They found that in western Albania the deformation was characterised by compression in a NE-SW direction at a velocity of 2.24 mm/yr. This is a similar magnitude to that observed in Montenegro, albeit with an azimuth of 58°, as opposed to 39° further north. In southwest Albania and northwest Greece, however, the deformation rate is 1.25 mm/yr (compression). The azimuth of the compression velocity vector here is 43°; near parallel to that observed along eastern Adriatic coast. In contrast with the compression in western Albania, in eastern Albania Papazachos *et al.* (1992) observe approximately 1.45 mm/yr of crustal extension, with a principal axis of extension oriented with an azimuth of 110°. This behaviour is maintained throughout western Macedonia and north-western Greece. The change in azimuth from 45° to 110° supports, in part, the model of rotation arising from the subduction of the Aegean plate within the Hellenic Arc, and is reflected across the tectonic regime of central Greece.

The contrast between a compressional regime in western Albania and an extensional regime in eastern Albania is further complicated by the presence of transverse fault zones across the country. The largest of these transverse fault zones is the Vlora-Elbasani-Dibra zone, which extends from the southern Adriatic coast of Albania to the Macedonia-Serbia border. This region has experienced many damaging earthquakes in its history, the most recent being the 1967 Dibra earthquake (M_W 6.6). The focal mechanism for this event (Baker *et al.*, 1997) suggests almost pure extensional faulting (though Muço (1994) identifies a small strike-slip component) and the isoseismals and surface fault trace confirm a NE-SW trending rupture plane (Sulstarova *et al.*, 2000). More detailed moment tensor analyses of earthquakes originating in this transverse fault zone show mostly dextral strike-slip motion, with a varying extensional or compressional component (Sulstarova *et al.*, 2000). This may represent the eastward extension of the boundary between the Adriatic microplate and the Africa plate, originally hypothesised by Anderson and Jackson (1987).

This pattern of compression along the western coasts and extension towards the east persists southward into northern Greece, only beginning to alter in the Ionian Islands. In the region of thrust faulting, focal mechanisms indicate a more north-westerly direction of strike. A clearer distinction between the two regions can be seen in the REIL2006 block model. Here another triple junction is defined just south of the Albanian border, from which a further boundary extends, running parallel to the northern Greek border. This margin is, again, poorly constrained by geodetic data and arises as part of the decision to define a new block in northern Greece. The strike of this margin is consistent with the focal mechanism of the Kozani-Grevena earthquake, though the comparatively high

sinistral strike-slip component of motion here is not reflected in the observed seismicity. The normal slip rate along this particular margin is similar to that of the parallel margin, which extends from the previous triple junction. In conjunction with the deformation rates estimated in Papazachos *et al.* (1992), this evidence suggests that the tectonic stress across in this region is roughly uniform in magnitude. If this is the case, the principal factor controlling long-term seismicity in the region is the orientation of the fault systems with respect to the direction of stress.

Focussing on the Adriatic region, there are many similarities in the HZ1985, PZ1990 and PP2000 source models. The coherent thrust fault structure that runs along the Adriatic coast clearly influences the delineation of source zones along the coast in all zone models. HZ1985 opts for the simplest solution, which is to model this as a single zone with a b-value of approximately 1.1. PZ1990 divides the region into three sub-zones (1a, 1b and 1c). This division is done on the basis of fault orientation. Zones 1a and 1b are assigned b-values of 0.8, whilst zone 1c is assigned a b-value of 1.0. The PP2000 model makes a further adjustment by dividing the Adriatic coast into 4 zones. Zones 1 and 4 appear to maintain the north-westerly orientation of the Adriatic-Eurasian thrust. The rationale behind the decision to separate western Albania into two source zones is not clear. The b-values for zones 1 to 4 all lay between 0.90 and 0.97, a result of the b-value smoothing described in Chapter 3 (Papazachos, 1999a). The difference in b-value between zones 2 and 3 is 0.03. Papaioannou and Papazachos (2000) do not give a full treatment of the assumptions behind their zonation. The only notable difference between the source parameters of zones 2 and 3 is in the azimuth of the major axis of isoseismals (-31° for zone 2 as opposed to -49° for zone 3). There is little variability in focal mechanism to support this assertion, and the few macroseismic intensities recorded in this region are unlikely to accurately constrain the axes of elliptical isoseismals. In short, there appears to be no obvious reason to distinguish between the two zones.

All of the existing seismogenic source models distinguish between thrust faulting on the Adriatic coast and the extensional faulting in eastern Albania and Macedonia. HZ1985 create a second zone (zone 6) that matches the shape of zone 1. This zone has a lower b-value (0.99) than that of the compressional zone. Again, the rationale behind the delineation of the eastern boundary of this zone is not clear. The neighbouring zone (16) has a substantially lower b-value (0.64), though the author notes that there is insufficient information from which to derive a b-value.

PZ1990 also distinguishes between the compressional and extensional fields. Here, two zones of compression (6a and 6b) in central Albania are shown. These are matched by two zones of extension (7a and 7b) in eastern Albania. Macedonia is left largely un-zoned. The decision to separate both these zones into an upper (a) and lower (b) sub-zone rests largely on the change in strike of the faults. Zone 7 is associated with a lower a- and b-values than zone 6, which is a reflection of the intraplate extensional nature of the region.

PP2000 opt to divide this same region into three zones (21, 22 and 23). Zones 21 and 22 are regular rectangles, zone 23 a trapezoid. The maximum magnitudes assigned to these zones are M_W 6.7, 6.6 and 6.4 respectively. Given the low upper limits on magnitude, the assumption of a point source is not unreasonable. However, in the case of zone 21 an earthquake of M_W 7.0 is believed to have occurred within the zone in 518 AD. The b-values for zones 21 to 23 lie between 0.89 and 0.93. As with the zones on the Adriatic coast, the only real distinguishing parameter between zones 21 - 23 is the azimuth of the major axis of isoseismals ($3°$, $16°$ and $-25°$ for zones 21, 22 and 23, respectively). The variability in focal mechanism makes the distinction between zone 21 and 22 insignificant, only zone 23 is appreciably different and this is largely reflected in the orientation of the zone.

Having seen how the tectonics of the Adriatic and Balkan region are represented in Aegean seismic source models, it is now possible to develop an original model. Within the Adriatic region, the vast majority of earthquakes, whose focal depths can be determined, originate in the upper 60 km of the crust. In keeping with the definition of shallow seismicity in the Aegean region by Papaioannou and Papazachos (2000), source zones in the Adriatic coast region can be considered homogenous to a depth of 60 km. The proposed zoning scheme for this region is shown in Figure A.3.

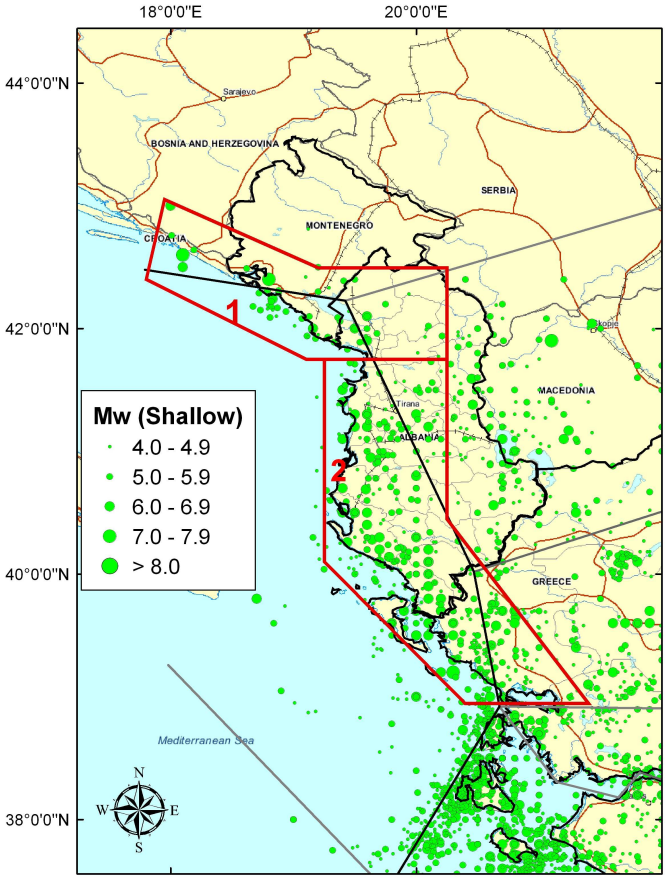

Figure A.3: Proposed zoning schemes for the Adriatic section of the Aegean region. The lines given on the maps refer to the faults proposed in the REIL2006 block model

The zonation scheme proposed is designed to capture two specific features of the seismicity in the Adriatic region. The first is the high seismicity attributed to thrust faults along the coast. The second is the difference between the compressional tectonics in the west of the region and the extension in the east.

Zone 1 includes the high seismicity around the Montenegro-Albania border. This region is extended northwest to include the strong earthquakes observed in the Montenegro-Croatia border region. The eastern terminus of this zone is located to include the triple junction inferred in REIL2006. For the Burton *et al.* (2004a) earthquake catalogue, b-values for this zone are 0.627 ($\pm$ 0.035) and 0.704 ($\pm$ 0.116) using least squares and maximum likelihood respectively. These values are quite low, and are most likely influenced by the inclusion of several large historical earthquakes.

The shape of zone 2 broadly follows a coherent line of large ($M_W > 6.0$) earthquakes. The width of this zone is to allow for uncertainty in the location of historical epicentres and for activity on smaller branch faults. Whilst orientation is an important consideration, it is ultimately the rate of slip that will have the greatest influence on seismicity. The comparable rates of slip for the two faults in Albania and northern Greece, inferred in REIL2006, support the delineation of one source zone.

The western boundary of zone 2 has been positioned to include as much of the seismicity that can be attributed to the fault system as possible. The southern boundary of this zone marks the transition from thrust faulting to strike-slip faulting in the northern Ionian Islands. This is positioned so that a small spatial cluster of powerful earthquakes is incorporated in to the seismicity of the zone. These earthquakes occurred in 1740 (M_W 6.2), 1867 (M_W 6.2) and 1898 (M_W 6.3). The b-values for this zone are 1.171 ± 0.094 (Least Squares) and 0.848 ± 0.050 (Maximum Likelihood).

A.2 Northern Greece, Macedonia and Bulgaria

The historical catalogue of earthquakes is only a sample of the true behaviour of seismicity. Over a longer sampling interval more events may appear in places not previously recognised as active. The risk in drawing uniform zones around clusters of powerful earthquakes is that it is implicitly assumed that a seismic zone will display stationary seismicity over time. There is always the possibility that a large unexpected earthquake will happen in a location not assigned as a seismic zone. In this case the zonation scheme would need to be reappraised after each event. Furthermore, if a zone is attributed on the basis of a single powerful observed earthquake, then the seismicity parameters are representative of a non-Poissonian process. The b-value may representative of the foreshock-mainshock-aftershock sequence of the large earthquake, not of the stationary seismicity of the zone. Alternatively, if the catalogues were declustered then there may remain too few events from which to determine seismicity parameters.

A good illustration of these problems can be found in the 1995 Kozani-Grevena earthquake (M_W 6.6). As noted by Stiros (1998), this earthquake occurred in a region deemed largely free of powerful earthquakes in previous hazard analysis, including the official seismic risk zoning of Greece published just days after the Kozani-Grevena earthquake. However, the historical and paleoseismic investigations conducted by Stiros (1998) and Karakaisis *et al.* (1998) indicate that the region has been struck by strong earthquakes (with maximum intensities of VIII or IX) with typical return periods of approximately 200 years. It is clear that the region cannot be considered aseismic. Nor can it be considered a single region of uniform seismicity either, as is assumed in the definition of a seismic source zone. A comprehensive investigation by Resor *et al.* (2005) has revealed that during this sequence, slip occurred across a much more complex network of faults than initially believed.

The region of northern Greece, Macedonia (specifically eastern Macedonia) and southern Bulgaria, needs to be considered separately from the seismicity of the Adriatic region, to the west, and the North Anatolian fault, to the south. This region is shown in Figure A.4, and is defined as the South Balkan extensional regime ((Burchfiel *et al.*, 2006). Within this region there are few, if any, obvious continuous linear fault structures. The observed seismicity is mostly diffuse (such as across north-eastern Greece and southern Bulgaria), but with some regions of spatial clustering (around Thessaloniki [A], Kozani-Grevena [B], the Gulf of Volos [C] and Struma [D]). Distinction is made between the cluster of seismicity in the Gulf of Volos [C] and the cluster in the Sporades basin. This is due to the abrupt change in focal mechanism between the normal faulting earthquakes across mainland Greece, and the strike-slip events at the western end of the North Anatolian fault.

The depths of earthquakes across this region are similar to those observed along the Adriatic coast. Hypocentres tend to originate within the top 60 km of the crust, with only a few outliers originating at any greater depth. This suggests that a conservative estimate for the depth of the seismogenic crust would be approximately 60 km, in keeping with the shallow seismic source definition of Papaioannou and Papazachos (2000). Further geophysical evidence to support this assertion can be found in the isodepth maps presented in Makropoulos and Burton (1984).

By considering the four clusters indicated (Kozani-Grevena, Thessaloniki, Gulf of Volos, Struma), it is possible to analyse the seismotectonic information available, to determine whether it is appropriate to designate these regions as uniform seismic zones. It has been shown that the 1995 Kozani-Grevena earthquake is not an isolated occurrence in the northern Greece (Macedonia) region (Stiros, 1998). Of the historical events in this region listed by Stiros (1998), only the 1995 event has a clear record of aftershocks. Attempts to determine b-value on the declustered catalogue of the Kozani-Grevena region are unreliable, as there are too few independent earthquakes in the catalogue from which to accurately determine the parameter. Of the 47 events identified in this region in the earth-

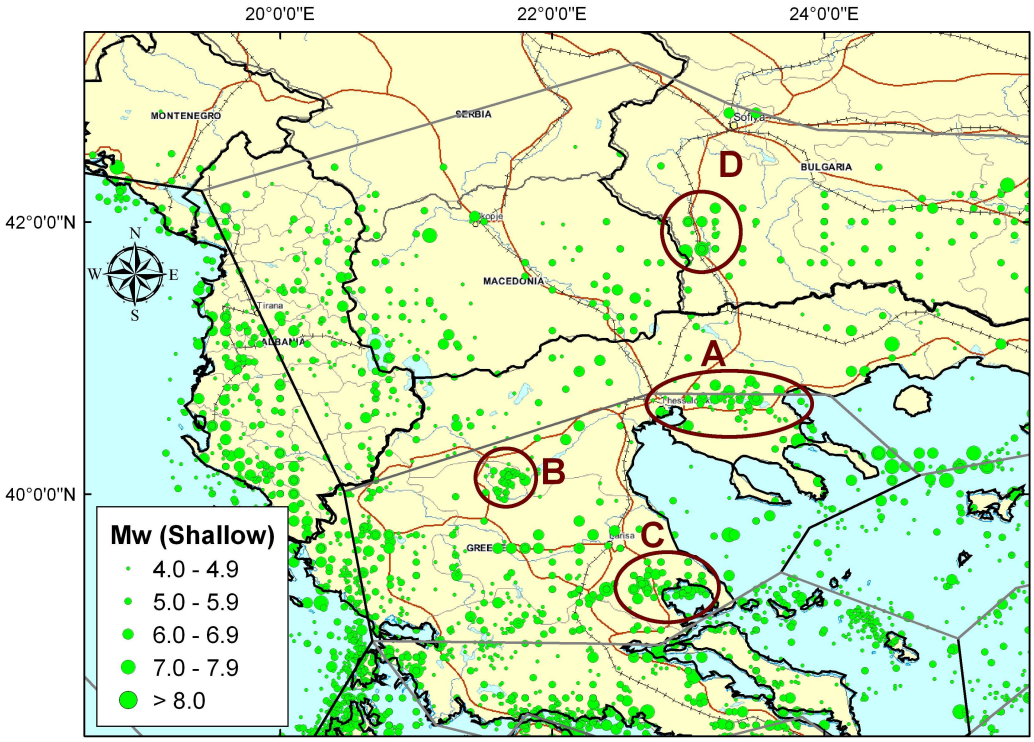

Figure A.4: Northern Greece-Macedonia-Bulgaria region, with earthquake hypocentres and REIL2006 block model shown.

quake catalogue, 41 occurred within three months of the May 1995 earthquake, with three more events post-1996. The b-value of the aftershock sequence is 1.23 (least squares), but the b-value for the zone using all the historical values is approximately 0.65. The influence of the two M_W 6.5 earthquakes in the catalogue is so great that a robust b-value cannot be assigned to the region. None of the existing source models for the Aegean isolate the Kozani-Grevena events.

Geodetic information for the Kozani-Grevena earthquake region is limited. Resor *et al.* (2005) use geodetic modelling to constrain the fault parameters of the earthquake sequence in this region, yet there is little mention of the long term slip. In the REIL2006 block model a continuous margin is defined, which runs SW-NE across the Kozani-Grevena region with 3.8 mm/yr of normal slip and 2.4 mm/yr of sinistral strike-slip. This is a small slip rate compared to that on the southern side of the block. GPS measurements within this region confirm that the slip rate is small compared to elsewhere in the Aegean. Strong earthquakes are clearly observed in the Kozani-Grevena region, and an active fault system is recognised, yet there is insufficient information to define a uniform seismic source zone.

It may not be appropriate to define the seismicity in the Gulf of Volos as an isolated cluster in the same manner as the Kozani-Grevena sequence. The proximity of this cluster to regions of active seismicity suggests a more explicit connection with seismogenic faults

in these regions. However, the E-W striking faults in this region become indistinct in the Pindos Mountains to the west (Goldsworthy *et al.*, 2002). There is a strong temptation to define a single continuous fault system to connect the crustal shortening in the Preveza region to the westward extension of the North Anatolian fault (NAF), yet there is little seismic evidence to support this.

The Gulf of Volos region has experienced a greater number of earthquakes than the Kozani-Grevena region. Whilst the seismicity is slightly more diffuse on a temporal scale, the historical catalogue is still dominated by time-dependent sequences of earthquakes around two events: 1957 Central Hills earthquake (M_W 6.7) and the 1980 Volos earthquake (M_W 6.5). This region has a b-value of 0.72 when declustered, as opposed to 0.92 without declustering. The latter value is consistent with the Papaioannou and Papazachos (2000) estimate for the region (0.89) and the decision in Papazachos (1990) to assign a b-value of 0.8 to the zone.

Geodetic observations for the Gulf of Volos region show a higher deformation rate than for much of northern Greece. REIL2006 interpret this as a block margin with dextral strike-slip of 13.9 mm/yr (similar to the movement on the NAF) and extension of 5 mm/yr. This rate is similar to the slip rate deduced by the seismic moment release (Papazachos *et al.*, 1992), of 9.5 mm/yr. It has already been established that the western trace of this fault system is unclear, and seismicity becomes more diffuse in the Hellenides. As this region is characterised by many highly active normal faults, it is consistent with the pattern of stress that exists for much of northern Greece. With no clear tectonic distinction between the seismicity to the west of the Hellenides and that around Thessalia, there is little reason to define any clusters as individual zones.

The third cluster under consideration is that observed to the east of Thessaloniki. The seismic activity of this cluster is attributed to the active Thessaloniki-Rentina Fault System (TRFS) (Tranos *et al.*, 2003). This is a coherent fault structure that strikes east-west from Thessaloniki to the Strymonikos Gulf (Figure A.5). Damaging earthquakes ($M_W \geq 6.5$) have occurred within this zone in historical times: 620 (M_W 7.0), 677 (M_W 6.5), 700 (M_W 6.6), 22 June 1759 (M_W 6.5), 5 July 1902 (M_W 6.6) and 20 June 1978 (M_W 6.5). Many of these earthquakes have caused damage (MSK $\geq$ VII) to the city of Thessaloniki, and analysis of the coulomb stress change following recent large earthquakes displays a westward progression; increasing the threat to Thessaloniki.

The Thessaloniki region is dominated by extension in the NNE-SSW direction, (Vamvakaris *et al.*, 2006) with faults dipping northwards. Focal mechanisms for the largest recorded earthquakes along the TRFS exhibit normal faulting, with some strike-slip events along the western section. The tension axes of the 1978 earthquakes are mostly oriented in the north-south direction (Papazachos *et al.*, 1991). Variation in these focal mechanisms is mostly due to the variation in strike of branch faults. In particular, the strike-slip earthquakes in the western section of the fault zone occur on faults strik-

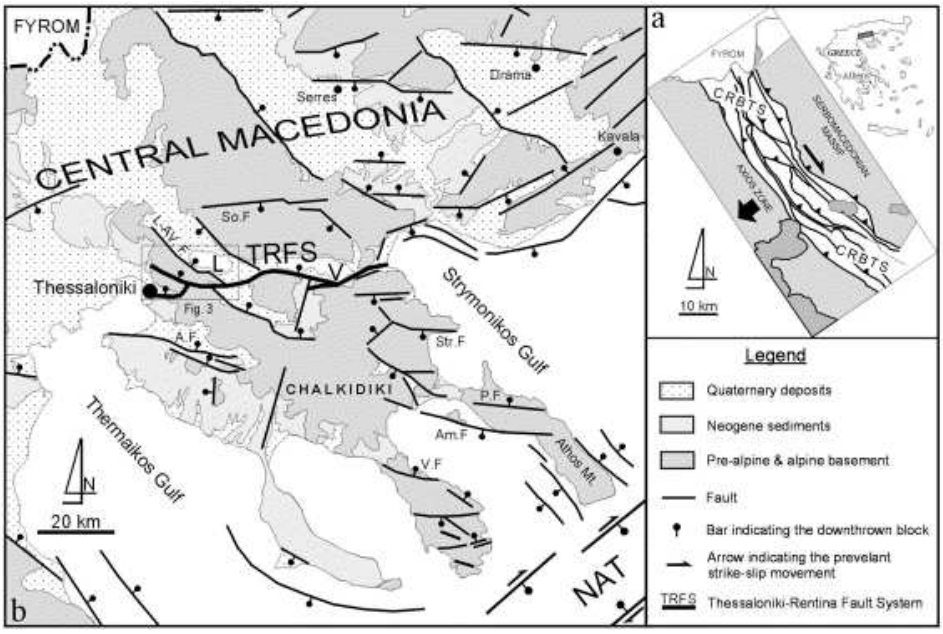

Figure A.5: Faults in the Thessaloniki Region (Tranos *et al.*, 2003)

ing NW-SE, which extend further northwest up the Langadas graben (Vamvakaris *et al.*, 2006).

The TRFS is recognised in many interpretations of Aegean geodynamics. The relatively small influence of the NAF on stress in this fault system is surprising, however. REIL2006 define the fault zone as a single block interface, striking east to west. They suggest 4.8 ± 0.7 mm/yr of sinistral strike slip, and 7.2 ± 0.6 mm/yr of normal slip. GPS vectors in the region define, relative to a fixed Eurasian frame, 1.3 ± 1.3 mm/yr of westward motion and 4.9 ± 1.2 mm/yr of southward motion (Burchfiel *et al.*, 2006), though this gives little indication of the slip across the fault itself.

There are some interesting distinctions in the way that the TRFS is incorporated into previous seismic source models. Both HZ1985 and PZ1990 define a single source zone running NW to SE across the Chalkidiki peninsular. This bears little relation to the observed strike of the fault zone, and must be influenced principally by observed seismicity parameters. This is maintained in the PP2000 model, with the definition of an additional source zone that covers the Thermaikos Gulf and a large part of northern Thessalia. It is not clear whether this definition arises as part of an attempt to define the southern Mygdonian basin as a single active source, or whether a connection to the NAF is the motive. Whatever the reason, it is difficult to see how a coherent, well-defined fault system, with observed seismicity and focal mechanism striking east-west, can be modelled as a zone of uniform seismicity orientated in a NW-SE direction.

Parameters of seismicity in the region are strongly influenced by the 1932/33 and 1978 events. The b-values are defined as 0.775 ± 0.049 (least squares) and 1.031 ± 0.171 (maximum likelihood) for the whole catalogue. If the catalogues is declustered, then the

b-values become 0.590 ± 0.053 (least squares) and 0.664 ± 0.130 (maximum likelihood). These values are generally consistent with the 0.6 to 0.8 range found in the existing source models.

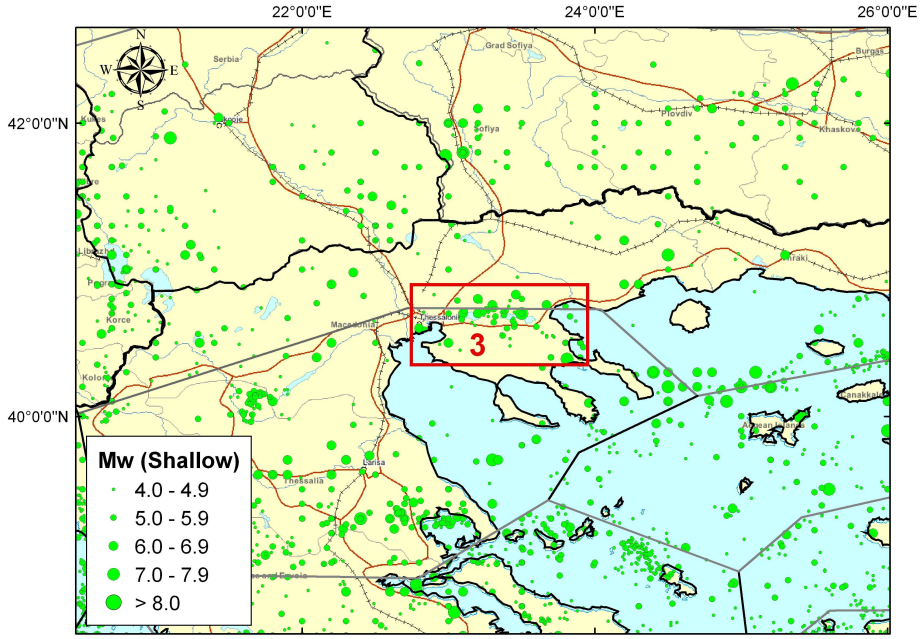

Figure A.6: Shallow source model proposed for the Thessaloniki region

Seismotectonic analysis of the TRFS indicates that it may be erroneous to define seismicity in this locality as an isolated cluster within the Greece and South Balkan area of diffuse seismic activity. Although focal mechanism and b-value are within the typical range of those found elsewhere within this region, the clear fault structure and high activity suggest that this is not an isolated cluster. Therefore, there is enough information to assign a zone of seismic uniformity to the Thessaloniki region. How, then, should the region of seismic uniformity be defined? The definitions seen in previous source models are not consistent with the dimensions of the observed fault system. Observed seismicity does not necessarily suggest a direct connection between the TRFS to the NAF, as implied in REIL2006.

Given the east-west strike of faults in the region, seismicity can be considered uniform within a small zone around the TRFS fault system. The epicentres define a simple line with a western terminus in the Thermaikos Gulf and an eastern terminus in the Strimonikos Gulf. Adding a 30 km buffer to this fault will incorporate the epicentres of the 1759 (M_W 6.5) earthquake and the 1677 (M_W 6.2) earthquake, both of which damaged Thessaloniki. Although neither of these events are within the TRFS defined by Tranos *et al.* (2003), their proximity to the fault system is such that their behaviour of these faults is highly likely to be linked to the TRFS via static stress transfer. This is transformed into a rectangular zone (Figure A.6). The eastern and western termini of the zone correspond to the ends of the E-W striking fault system. Northern and southern boundaries are determined using the epicentral distribution of powerful earthquakes that have been

observed along the TRFZ.

A.3 Ionian Sea and the Cephalonia Transform Fault

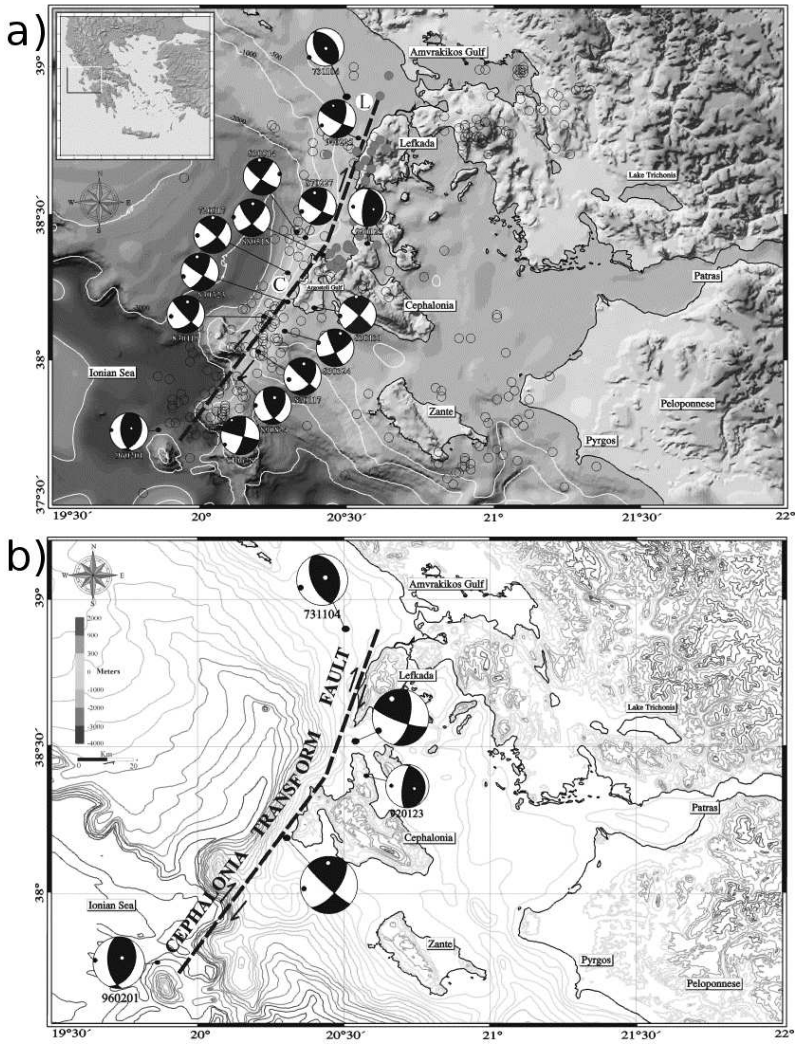

Figure A.7: Location of the Cephalonia transform fault (Cephalonia and Lefkada segment) (Louvari *et al.*, 1999). a) Focal mechanisms of recorded strong events, b) "Typical Focal Mechanism" for each segment. Note the north and south termini of the fault system are indicated by the locations of the thrust faulting earthquakes of 11 November 1973 and 1 February 1996, respectively.

The continental collision between the African plate (and connected microplates) and Eurasia (via the Aegean and Anatolian microplates), is clearly visible in the thrust faulting observed along the Hellenic Arc and Adriatic coast. Nevertheless, there is a clear distinction between the subduction observed along the Hellenic Arc, and the orogenic collision that forms the Dinarides and Albanides further north in the Balkans. This transition is marked by a shorter line of highly active dextral strike-slip seismicity, which extends a total length of approximately 125 km (Louvari *et al.*, 1999). The southern tip of this fault system is located in the Ionian Sea approximately 70 km west of Zante, and

extending northeast before terminating approximately 10 km north of Lefkada (Figure A.7). This fault system is divided into two segments: the Cephalonia segment (85 km length, 20 km width) and the Lefkada segment (40 km length, 15 km width). The two segments are distinguishable largely by strike, the Cephalonia segment striking at approximately $38°$, and the Lefkada segment at approximately $14°$. The transition between the two segments occurs in the Myrtos Gulf in northern Cephalonia (Figure A.7 - Louvari *et al.* (1999)).

Despite being a small fault system in comparison to the Hellenic arc and NAF, much more slip is observed here. This produces the high seismic activity. Estimates of the dextral slip rates range from 14.5 mm/yr (Papazachos *et al.*, 1992) to 41.3 mm/yr (Reilinger *et al.*, 2006). Most estimates seem to fall within the middle of this range. The slip estimate of approximately 30 mm/yr (Papazachos and Kiratzi, 1996) is in good agreement with the GPS velocity of estimate 28 ± 3 mm/yr (Cocard *et al.*, 1999; Nyst and Thatcher, 2004). This implies that almost all of the slip on the fault is coseismic.

The Cephalonia segment of the fault has ruptured 8 times in the last 600 years to produce earthquakes with $M_W \geq 7.0$. The largest earthquake to have occurred on the segment is the 4 February 1867 event (M_W 7.4). Two events of comparable magnitude have occurred in the 20th century: 12 August 1953 (M_W 7.1) and 17 January 1983 (M_W 7.0). Whilst the Lefkada segment would appear to be too short to accommodate enough slip to produce $M_W > 7.0$ earthquakes, $M_W > 6.0$ events occur with greater regularity. These include the 27 November 1914 (M_W 6.2), 22 April and 30 June 1948 (M_W 6.5 and M_W 6.4 respectively), 4 November 1973 (M_W 6.0) and 14 August 2003 (M_W 6.2) events (Karacostas *et al.*, 2004). These events caused damage on the island of Lefkada on many occasions. Four events with $M_W \geq 6.7$ have occurred on the Lefkada segment in the historical catalogue, which corresponds to the approximate maximum magnitude estimated from the length of the segment (Wells and Coppersmith, 1994).

When creating seismic zones in the region of the Ionian Islands one of the first decisions that should be made is whether to consider the Cephalonia and Lefkada segments as separate sources. The HZ1985 and PZ1990 models consider the region of strike-slip faulting as one single seismic source, whereas PP2000 separate it into a Cephalonia zone and Lefkada zone. Inherent within this question is the issue of whether seismicity on the Lefkada segment can be considered independent of that on the Cephalonia segment.

An initial inclination would be to separate the Cephalonia and Lefkada segments because of the difference in source dimensions, and therefore maximum magnitudes. This does not immediately suggest that the segments are independent. Studies of static stress change following the 2003 Lefkada earthquake (Karacostas *et al.*, 2004; Papadimitriou *et al.*, 2006) indicate that stress can be transferred from one segment to another. Given the comparatively small differences in strike and dip between these two segments, it is likely that this stress change will result in periods of high activity along the neighbouring

segment. Karacostas *et al.* (2004) speculate that the 27 January (M_W 6.6) and 7 August (M_W 6.7) 1915 earthquakes on the Cephalonia fault segment may have been triggered by an M_W 6.3 event the previous year on the Lefkada segment. More recently, coulomb stress change calculations following the 2003 Lefkada earthquake identify an increase in coulomb stress at the northern end of the Cephalonia segment (Papadimitriou *et al.*, 2006). A cluster of small earthquakes, including one M_W 5.1 earthquake three months after the Lefkada event, were situated in the region of highest positive coulomb stress change.

The identification of time-dependence between the two segments by coulomb stress transfer suggests that the Lefkada and Cephalonia segment of the transform fault cannot be treated independently. It is therefore appropriate to designate the entire transform fault system as one zone. Whilst the difference in strike may have implications for hazard analysis when considering directivity effects, this is small compared to the uncertainties that emerge in modelling ground motion directivity.

The region immediately south of the Ionian Islands is also seismically active. This represents the western extension of the Hellenic Arc, where intermediate depth earthquakes begin to emerge. Here, shallow seismicity displays predominantly thrust faulting. The transition between transform faulting and thrust faulting is quite abrupt and can be seen in focal mechanisms for strong earthquakes in the region. The hypocentral distribution shows no such abrupt change and remains well dispersed over much of the southern Ionian region. This presents problems when defining a zone to characterise seismicity that is related to, but not directly attributable to movement on the CTF.

The choice was made to encapsulate the remaining seismicity (i.e. that not directly attributable to the CTF) in the Ionian Islands in one single uniform zone (Figure A.8). The zone encroaches into south-western Greece. This is to include some of the mixed thrust and strike-slip faulting seen in this area, which is more likely to be influenced by activity along the Hellenic Arc, rather than the extensional faulting that is prevalent across much of southern Greece.

Proceeding further southeast along the Hellenic Arc, the influence of the subduction zone increases and that of the CTF diminishes. Focal mechanisms in this region are similar to those running all along the Hellenic arc, which offers little indication as to where the Ionian zone should terminate. GPS estimates of crustal velocity (Nyst and Thatcher, 2004; Hollenstein *et al.*, 2006) indicate that deformation is uniform across this region. The southeastern boundary therefore transects a region of lower seismic activity between the southern Ionian Islands and the southern coast of Greece. It runs immediately to the northwest of the epicentre of the 27 August 1886 Peloponnese earthquake (M_W 7.5) and nearby cluster.

The CTF zone (4) has b-values of 0.907 ± 0.046 (least squares) and 1.004 ± 0.079 (maxi-

mum likelihood). The larger Ionian zone (5) has b-values of 1.053 ± 0.061 (least squares) and 1.225 ± 0.079 (maximum likelihood). This is similar to the estimates of HZ1985 for the Ionian region who give b as 1.03. PZ1990 assign a b-value of 1.0 to the region, whilst PP2000 suggest a value of 0.99 for all the zones incorporated into the Ionian region. The frequency of strong earthquakes attributed to the CTF would appear to account for the lower b-value in this zone.

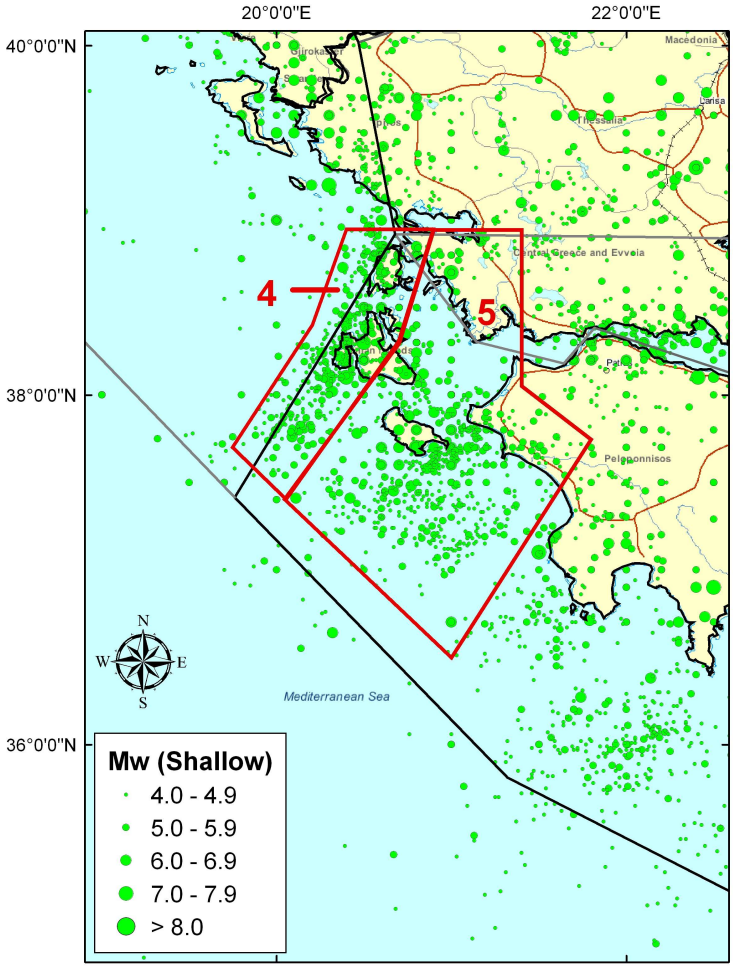

Figure A.8: Proposed shallow source zonation for the Ionian Islands

A.4 Central Greece and the Gulf of Corinth

The Gulf of Corinth is recognised as being the most seismically active region of continental Greece. Although historical earthquakes have exhibited lower magnitudes than those in the Hellenic Arc, CTF and NAF, they have been no less destructive. This is partly due to the shallow nature of the faulting, the abundance of soft soils, and the high number of large conurbations around the Gulf of Corinth. Active offshore faults mean that tsunami hazard is also present in this region.

The Gulf of Corinth is identified as an asymmetric rift, with the most active faults dip-

ping north (Bernard *et al.*, 2006). Estimates of geodetic movement across the entire Gulf suggest horizontal movement rates between 11mm/yr and 16 mm/yr. It displays a substantial amount of internal deformation, in some places greater than 120 nanostrain/yr (Avallone *et al.*, 2004). This deformation is not consistent across the entire Gulf, however. Analysis of GPS data by Clarke *et al.* (1997) found that GPS velocities were larger in the west of the Gulf (13 ± 1 mm/yr) than in the east of the Gulf (6 ± 1 mm/yr). This result is borne out by later observations (Moretti *et al.*, 2003; Avallone *et al.*, 2004; Nyst and Thatcher, 2004).

It is because of the significant seismic hazard in the Gulf of Corinth, in addition to its geodynamic importance, that such resources have been directed into researching the area. The dimensions of many active and inactive faults are well constrained by geophysical data, both on and offshore (Figure A.9). Extensive paleoseismological research has been undertaken in the Gulf of Corinth (Chatzipetros *et al.*, 2005) and continual monitoring of active faults in the western Gulf of Corinth has been undertaken by the Corinth Rift Laboratory (Bernard *et al.*, 2006).

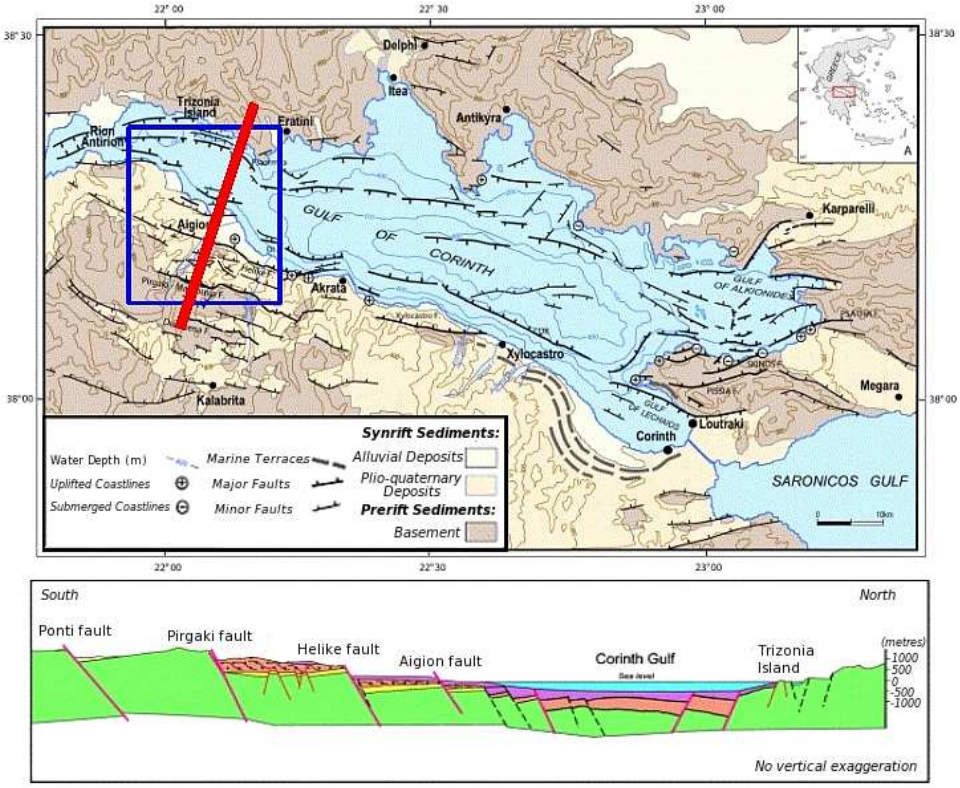

Figure A.9: Active faults in the Gulf of Corinth region (upper) with cross section along the red line shown (lower) (Moretti *et al.*, 2003). Blue square indicates the location of the 3D section shown in Figure A.10

Figure A.9 indicates that faulting in the Gulf of Corinth is in a WNW-ESE direction. Focal mechanisms for the Gulf of Corinth are in agreement with this observation. The primary axes of tension strike in a N-S direction, which is consistent with the extensional nature of mainland Greece. A number of hypotheses exist as to why this rift is formed, the

favoured one being a combination of lithospheric thinning as part of a back-arc basin, in addition to gravitational collapse of a thick crust that arose from Neogene mountain building (Moretti *et al.*, 2003).

Although the Corinth region displays largely extensional faulting, there are large differences in the seismicity observed across this region. The first notable feature is the contrast between the high seismicity in the western Gulf of Corinth and the low activity immediately to the west in the Gulf of Patras. The western Gulf of Corinth appears to have some of the highest seismicity in the region. It is from this location that the Corinth Rift Laboratory operates, and the seismogenic behaviour of these faults is well known. There have been many large historical earthquakes in this region, five with $M_W > 5.8$ in the last 35 years, the most recent being the 1995 Aigion earthquake (M_W 6.2). A detailed analysis of faulting around the Aigion region (along the southern coast of the Gulf of Corinth) is seen in Bernard *et al.* (2006) - Figure A.10. It is apparent that there are a greater number of seismogenic faults along the southern coast of the Gulf. Some of the large historical earthquakes (including the 373 BC and 1861 AD events) have been attributed to the Elike fault (Stewart, 1996). Recent seismicity data indicate that the western section of this fault is currently locked. This has prompted the suggestion that it may have even been deactivated by the emergence of the younger Kamarai fault system (Bernard *et al.*, 2006).

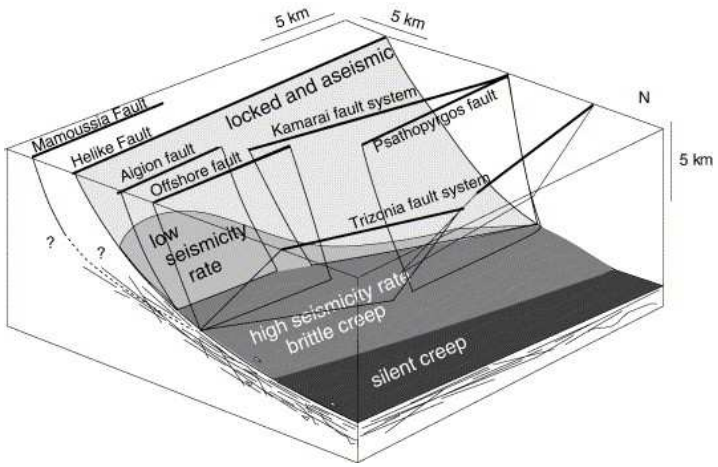

Figure A.10: 3D section through the faults along the south western coast of the Gulf of Corinth. Model based on data from seismic swarm in 2000-2001 (Bernard *et al.*, 2006)

The existence of such a wealth of geophysical data may assist in characterisation of a source zone in this region, but it is still far from conclusive. It is not yet clear how these faults interact and if they display interdependence. In particular, the question remains as to the nature of the link between seismicity in the western Gulf of Corinth, and that in the east.

The tectonics of the eastern Gulf of Corinth and the eastern coast of Greece are no less complicated than those to the west. As with the rest of mainland Greece, it is extension that characterises much of the region. Here there are many graben-like structures, both

within the Gulf itself and across the eastern coast. The eastern Gulf of Corinth divides in to two basins: the Alkyonides Gulf and the Gulf of Lechaios (Moretti *et al.*, 2003). To the northeast of the Gulf of Corinth another graben-like structure exists on the eastern coast of mainland Greece: the Gulf of Evoikos. Whilst many comparisons between the Corinth and Evoikos basins may be drawn, the western extension of the NAF has a significant influence on the deformation and seismicity in the Evoikos Gulf, which is not apparent in the Corinth Gulf (Papoulia *et al.*, 2006).

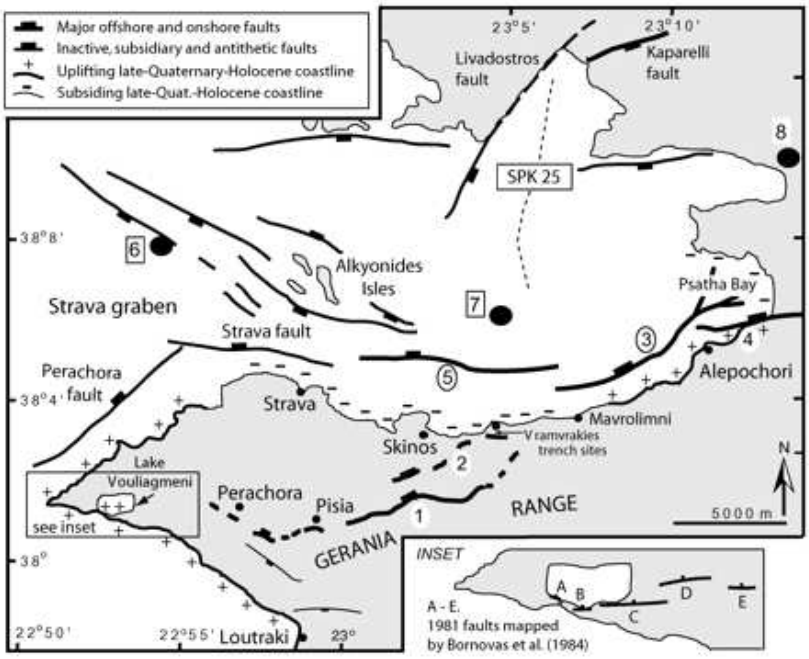

Figure A.11: Faults in the eastern Gulf of Corinth [Alkyonides] with normal faulting in the quaternary period (Leeder *et al.*, 2005)

As with the western Gulf of Corinth, there exists a complex network of active faults throughout the Alkyonides and Lechaios region (Figure A.11). Most of these faults strike in an E-W direction, displaying clear normal faulting with extension in the N-S direction. This is confirmed in the focal mechanisms of Corinth earthquakes found in Kiratzi and Louvari (2003). Further east the typical strike changes from WNW-ESE to more pure E-W. Although the rate of opening in this section of the Gulf of Corinth is lower than that observed further west (Clarke *et al.*, 1997), the cumulative amount of tectonic extension would appear to be greater in the east. This is clearly visible in the geomorphology and bathymetry of the eastern Gulf, which is substantially wider and deeper than in the Aigion region to the west. This provides a conundrum when trying to account for all the geodetic slip using the cumulative tectonic moment rate. Clarke *et al.* (1997) addressed this problem, suggesting that this moment deficit in the Western Gulf could be accommodated by large earthquakes in the future. There is very little other information to confirm this hypothesis and the impact of aseismic creep or static stress transfer has yet to be determined.

Beyond the eastern extent of the Gulf of Corinth, into the Attica region and Gulf of Saronikos, more variation in seismicity and focal mechanism is visible. This region is of particular importance because of the potential hazard to the city of Athens. Since the 7 September 1999 earthquake (M_W 5.9) it has become more prominent in seismic research in Greece. As with the Gulf of Corinth (east) and the Gulf of Evia (North), this region is dominated by extensional faulting. The focal mechanism of the 1999 Athens earthquake shows extensional faulting striking at 115° (Kiratzi and Louvari, 2003).

The strikes of normal faults in the Attica region do exhibit some variation compared to those in the eastern Gulf of Corinth. There exists a coherent fault structure linking the Alkyonides Gulf to the Gulf of Evia, which is referred to as the Kaparelli-Oropos fault zone (Ganas *et al.*, 2004). Many large earthquakes have been recorded in this zone, the most recent being the 4 March 1981 (M_W 6.4) event. Along the zone, faults seem to strike in a WSW-ENE direction ($60^\circ - 70^\circ$), which contrasts to the WNW-ESE striking faults in the Mount Parnitha region.

There are very few large earthquakes in the Gulf of Saronikos region for which focal mechanisms have been constructed. The situation is complicated further by a number of volcanic and geothermal systems, including those on the island of Methana and Aegina, which produce localised variations in stress. Because of the seismic hazard the Saronikos region poses to the city of Athens (Burton *et al.*, 2004b) caution must be exercised in assuming that this low seismicity is persistent.

Geodetic analysis of slip and deformation in the Attica region indicates that the extensional movement occurs at a substantially lower rate than in the Gulf of Corinth. Measurements of mean slip rates across active faults in the Mount Parnitha region indicate slip of 0.18 - 0.25 mm/yr (Ganas *et al.*, 2004). The authors suggest that this result would manifest itself seismically in the form of large earthquakes with long recurrence intervals. The rates for the Parnitha region are comparable to the 0.2 - 0.4 mm/yr slip estimated for the Gulf of Evia region. This is an order of magnitude lower than the slip rates estimated for the Gulf of Corinth.

There is abundant evidence of faulting in the Gulf of Evia during the quaternary, yet historical seismicity is largely inactive, manifesting as clusters of microseismicity and sporadic large events. One of the more curious structures is the Atalanti fault, which ruptured in 1894 to produce an especially destructive earthquake (M_W 7.2), but today exhibits seismic inactivity. Jackson (1999) interprets the comparative behaviour of the Evia and Corinth gulfs as an example of "fault death". This implies that pre-historic extension in the Evia graben has been transferred into the extensional regime in the Gulf of Corinth. Exactly why this transfer should occur is a matter of speculation, with possibilities being the rotation of the block to a position unfavourably oriented for activation, stress interaction with nearby faults (particularly the NAF), or strain-hardening.

Before considering how to zone the Central Greece region, reference is once again made to the REIL2006 model. This defines a "Central Greece" block with faults bisecting the Patras and Corinth gulf, turning northeast along the Kaparelli-Oropos fault zone into the Gulf of Evia, before turning southeast along the Evia Gulf and into the Aegean. The northern margin of the block strikes due east from the CTF to the Sporades islands. As established previously in Section A.2, there is no evidence from the observed seismicity to define a continuous structure in this region.

Previous seismic source models have all separated the Gulf of Corinth region from the rest of Greece and the Hellenic Arc. HZ1985 has a single continuous zone running E-W across the Gulf of Corinth before turning southeast into the Saronikos Gulf. No eastern boundary is defined, so the zone expands into the south Aegean. The Gulf of Evia is defined as another zone, again unbounded, running NW-SE into the Aegean. There is a striking contrast in the b-values for these two zones that are presented in HZ1985, with the Gulf of Corinth zone having a b-value of 1.01, compared to 0.86 for the Evia region. The higher b-value for the Corinth region is likely due to the higher rate of seismicity observed in the gulf.

The PZ1990 model separates this part of central Greece into four zones: three along the Gulf of Corinth and one in the Gulf of Evia. The division of the Corinth Gulf into three zones is based on the contrast between the seismicity in the east and west of the Gulf, with the third zone covering the Saronikos Gulf (required because of the lower seismicity rate). Where this model is flawed is that it leaves much of the Attica region un-zoned, a shortcoming that was highlighted in the Athens earthquake, which occurred from an unexpected source. The b-values for these zones are largely identical to those of HZ1985, with all three Corinth zones having b-values of 1, and the Evia zone 0.8.

The source model of PP2000 is similar to the PZ1990 model, albeit with further subdivisions. Here the Gulf of Patras is encapsulated within a single zone, and the Gulf of Corinth separated into an east and west section. The Gulf of Saronikos is defined as a single zone, and the Evia region split into a north and south zone. The south-eastern Attica region is encompassed in the low-seismicity South Euboikos zone, which covers a portion of the South Aegean. There is little variation in b-value across all of the zones in this region with the two Evia zones and the South Euboikos zone all having a b-values of 0.90, though the a-value and M_{MAX} for the South Euboikos zone are lower. The Gulf of Corinth and Saronikos zones also have similar b-values (0.93 and 0.92 respectively) with only the Patras zone having anything higher (b = 0.96). It is not unreasonable to suggest that all the b-values for the central Greece region are within the margin of error that in estimation of the Gutenberg-Richter parameters, as discussed in Chapter 2.

One of the most contentious issues when creating a seismic source zone for central Greece is whether to divide the seismicity in the Gulf of Corinth. Both PZ1990 and PP2000 split the zone arbitrarily in the middle of the Gulf of Corinth, to separate the seismicity in

the east from that in the west. The problem that arises is that this division bisects the Xylocastro fault: a coherent and active system that connects the two regions of seismicity in the Gulf. This sheds some doubt on the validity of the assumption that the two regions of seismicity are independent. The issue is complicated yet further by the evidence of Cowie and Roberts (2001) that Xylocastro fault is the central fault of a system of growing normal faults; the central fault slipping at three times the rate of its distal sections (the distal sections being the Aigion and South Alkyonides faults).

Although further research is required into the time-dependent nature of faulting in the Gulf of Corinth, division of this region into any separate zone system may be erroneous. Furthermore, where an attempt to make a distinction between the eastern and western Corinth regions is undertaken (PZ1990 and PP2000), the parameters of the respective sub-zones are so similar as to make distinction practically pointless from a hazard perspective. It is because of this that in this model only a single zone is created, one that covers the entire Gulf of Corinth (Figure A.12). The zone presented in this source model is designed to incorporate as much of the seismicity resulting from the high rate of extension in central Greece as possible. Comparison can be made with the HZ1985 and EK1999 models of seismic source zones. The decision to terminate the western boundary of the zone in the Gulf of Patras is so as not to impinge on the strike slip faulting seen closer to the Ionian Islands (already captured by a previous zone).

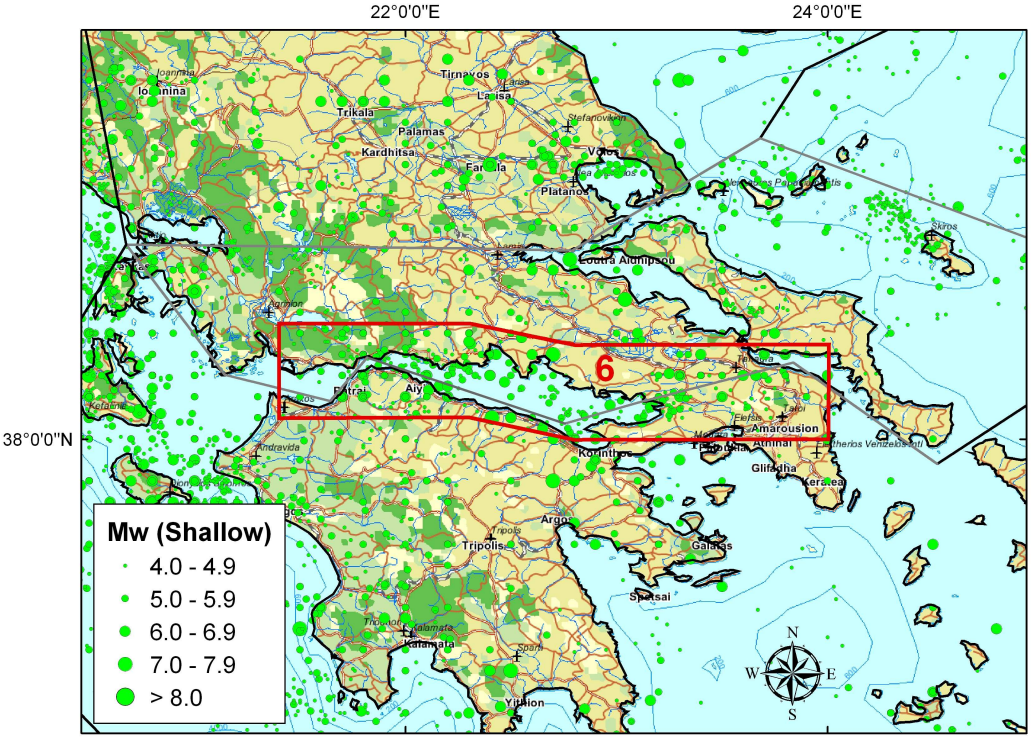

Figure A.12: Proposed zonation for the Gulf of Corinth Region

It has been the decision of previous source modellers to extend the Corinth system into the Gulf of Saronikos, whilst separating the Evia region into a different source zone. The

only exception to this is the EK1999 model which ascribes a triangular zone encompassing all the graben-like features in the Central Greece extensional zone. Both approaches are avoided here. The low activity and dispersed seismicity in the Saronikos Gulf suggests that any attempt to characterise it as uniform, let alone directly connected to that of the Corinth Gulf, would be misguided. Similarly, whilst the Gulf of Evia conforms to a similar pattern of extension, the low slip rates mean that it cannot be reasonably considered part of the Corinth system. In both regions earthquakes seem to originate from complex structures whose characteristics are still poorly defined. In that respect they cannot be considered uniform in any sense. The only information available to characterise these regions is the observed seismicity. Hence, both the Gulf of Evia and Saronikos are excluded from the uniform Corinth zone.

The eastern terminus of the Corinth zone is located to incorporate as much of the seismicity in the Attica region as possible into the Corinth zone. The occurrence of damaging earthquakes (and the potential hazard to Athens) in this region mean a greater attempt must be made to characterise the hazard here. Although there is some distinction between the strike angles of faults in the Attica region and those in the Gulf of Corinth, it is not sufficiently large to have an impact on directivity effects when modelling hazard, given the large uncertainties in source parameters. As such, it is not inaccurate to attribute seismicity in the Attica region to a similar style of faulting as that of the Gulf of Corinth.

The b-values for the zone proposed are 0.864 ± 0.064 (least squares) and 0.909 ± 0.073 (maximum likelihood). This is within the estimated range from the previous models, the closest being the HZ1985 model. The maximum observed magnitude over the 2000 year historical catalogue is 7.0.

A.5 The Hellenic Arc and South Aegean Sea

The Hellenic Arc is a subduction zone that extends from the Ionian Islands to Western Turkey. It is the clearest demarcation of the boundaries of the African plate and the Aegean plates. The subduction of the African lithosphere is the principal mechanism that drives the tectonics of much of the Aegean and Balkan region. Seismically, this region is subject to high activity with the most powerful earthquakes in the historical catalogue (three with $M_W \geq 8.0$) being recorded along sections of the arc. It is also in the South Aegean and Cretan Sea region that some of the deepest seismicity of the Eastern Mediterranean is seen. While intermediate depth (60 km - 200 km) seismicity may not be associated with as high a hazard as shallow seismicity, its contribution to hazard at a site is not necessarily trivial either. This is especially true when considering large engineered structures, which respond more radically to long period oscillation. The Hellenic arc may be the most seismically active region of Europe, but not compared to other subduction zones across the globe (Bohnhoff *et al.*, 2005).

Although some intermediate depth seismicity can be seen elsewhere in the Aegean, this has often been in the form of isolated events. Beneath the Cretan and South Aegean Sea, the intermediate depth seismicity defines the Benioff zone of the down-dipping slab (Papazachos, 1990). It is because of this that a two layer seismic source model is used. The upper layer is defined in the same manner as for previous zonations in the Aegean, using shallow (depth $<$ 60 km) seismicity, and it can be considered homogenous to that same depth. The lower layers refer to intermediate seismicity (depth 60 - 180 km).

The rate of convergence between the African and Eurasian plate systems in the Hellenic Arc is approximately 40 mm/yr (McClusky *et al.*, 2000). Most of this is accommodated as compression along the Hellenic thrust itself. Focal mechanisms reveal a more complicated situation. Along the convex side of the arc most of the well-defined fault plane solutions indicate NW-SE strike, dipping between 20° and 30° (Papazachos *et al.*, 1984). Fault plane solutions for plate boundary events south of Crete confirm this trend, accurately constraining the principal axis of pressure to a SSE (N 190° E $-$ N 220° E) direction, with dips close to 20° in a northeast direction (Bohnhoff *et al.*, 2005). These are the typical focal mechanisms that define most of the western outer Hellenic Arc.

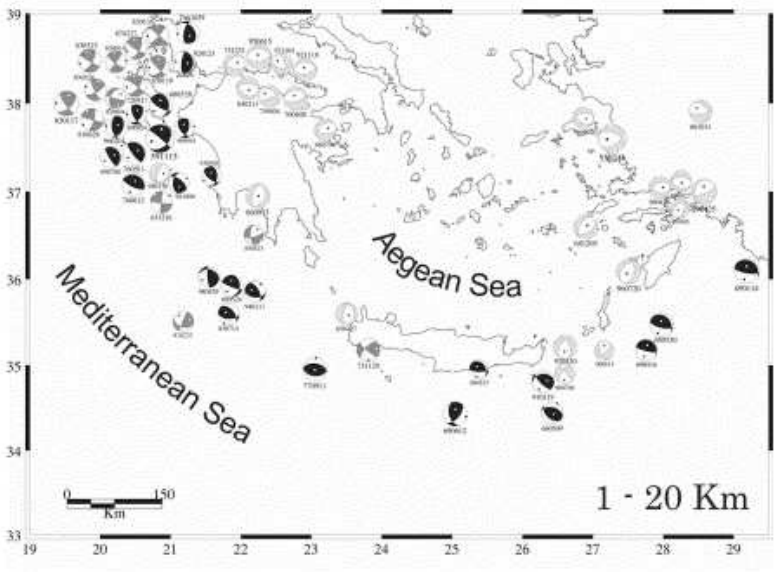

Figure A.13: Focal Mechanisms for shallow earthquakes in the Hellenic Arc (Benetatos *et al.*, 2004)

Although the main trace of the Hellenic trench is visible in the bathymetry as far as Rhodes and western Turkey, earthquake focal mechanisms still indicate thrust faulting (Figure A.13). The change of direction of the margin from NW-SE to SW-NE results in a more substantial strike-slip component of faulting. Strike-slip earthquakes in this section (Bohnhoff *et al.*, 2005) of the arc are often isolated, resulting from localised variations in stress.

The inner [concave] side of the Hellenic Arc displays an entirely different faulting mechanism (Figure A.13). In this region focal mechanisms and fault scarps reveal predom-

inantly normal faulting, with principal axes of tension striking in an E-W direction (Armijo *et al.*, 1992). This trend is visible in the shallow seismicity, and runs parallel to the main thrust of the arc. The geology of this zone indicates that it is an accretionary prism that leads up to the volcanic arc (Benetatos *et al.*, 2004). Exactly why this arc-parallel extension is observed is still a matter of some speculation. Armijo *et al.* (1992) propose that this extension is a result of the subduction of increasingly buoyant lithosphere, as the continental margin of the African plate is pulled closer to the Aegean. This hypothesis is supported by the observation that the arc-parallel zone of extension does not extend that far into the South Aegean Sea (or to significant depth). This suggests that the Aegean lithosphere is weak, an inference made in earlier work by McKenzie (1978).

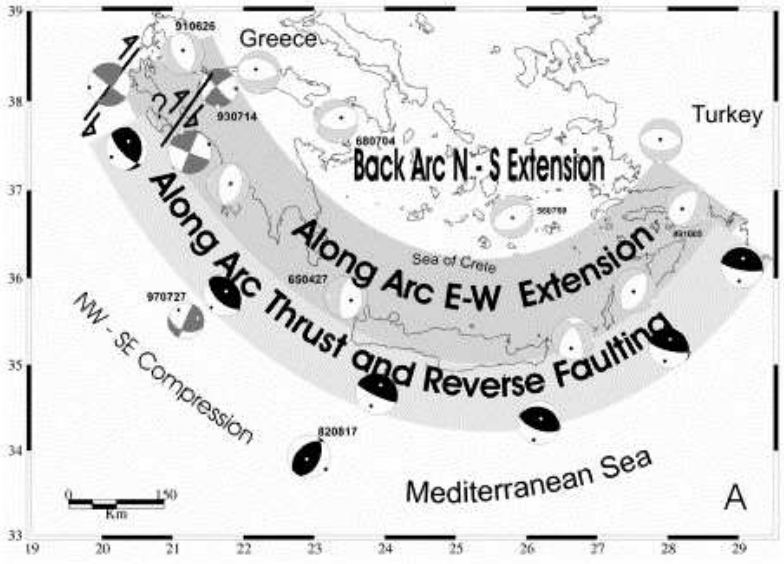

Figure A.14: Interpretation of focal mechanism data from Benetatos *et al.* (2004)

The south Aegean Sea and volcanic arc is an area of lower seismicity. Focal mechanisms in this region indicate extension in a N-S direction. This is consistent with the model of a back-arc basin that extends from the Peloponnese, across the Aegean Sea and well into western Turkey. This area is also associated with line active of volcanism, with four volcanoes (Methana, Milos, Santorini and Nisyros) having erupted since human settlement in the area (Luhr, 2006).

Although much of the south Aegean sea displays low seismicity, there is a thin band of active normal faulting that originates in western Turkey and terminates in the southern Cyclades islands (around the island of Santorini). Focal mechanisms in this region are well aligned with the N-S extension, yet activity is significantly higher than elsewhere in the south Aegean Sea. This is a feature often overlooked in tectonic models of the Aegean region.

For the Hellenic Arc region, the REIL2006 geodetic model presents some curiosities. The plate boundary is well-marked by a series of connected thrust faults running from the Ionian Islands to eastern Crete. In the Dodecanese islands, the style of movement along

this margin changes from thrust to almost pure strike-slip. This transition would appear to be logical given the relative movement of the African and Eurasian plate, which is why it is strange that such a transition is not apparent in the focal mechanism. Along the arc the strike-slip component of movement changes from dextral, in the western Hellenic arc, to sinistral in the eastern arc. Again, this transition is also consistent with the broader geodynamic framework of the eastern Mediterranean. Observed slip rates along the arc are consistent with GPS estimates of the Africa-Eurasian collision, with approximately 40 mm/yr total slip along the western arc, rising to nearer 50 mm/yr along the eastern arc.

The main curiosity in the REIL2006 model is the inclusion of an eastern Hellenic block. This is a comparatively small block bounded by the eastern Hellenic thrust to the south and east, and a perpendicular set of extensional oblique faults to the north and west. The northern boundary of this block runs along the line of high seismicity that extends from western Turkey to the Cyclades islands. This boundary has a comparatively high rate of extensional-oblique motion, with 10 - 20 mm/yr extension and approximately 8 mm/yr sinistral strike-slip motion. The western boundary strikes mostly N-S, almost perfectly bisecting the island of Crete. Movement along the western boundary is a composite of E-W extension and N-S dextral strike-slip motion.

The definition of an east Hellenic block is unprecedented in previous kinematic models of the Aegean (Nyst and Thatcher (2004) - and references within). The correlation of the northern boundary of the block to the band of high seismicity (be it intentional or accidental) provides an interesting explanation for the seismicity, and why it should terminate so abruptly in the Cyclades. There is little seismic or geophysical evidence for a coherent fault system bisecting Crete, however. It could be speculated as to whether the definition of a N-S striking line of extension is merely a convenient kinematic way of accommodating the E-W extensional strain that is seen across the concave side of the Hellenic Arc. If this is so, then this particular block model may not be of any substantial use in the development of a seismic source model.

There is some disparity in the way in which the Hellenic Arc has been defined in previous source models. The EK1999 model defines a single zone that matches the shape of the arc, extending from the Ionian Islands to the western Dodecanese islands. The remainder of the Dodecanese islands are encapsulated within a single heptagonal zone. Furthermore, the Saronikos-Cyclades-Dodecanese arc is defined as a single zone. The designation of a single zone of seismicity in the Peloponnese is more consistent with observations.

The HZ1985 model has some similar definitions to the EK1999 model in that the Hellenic Arc is separated from the Peloponnese, Dodecanese and Cyclades zone of seismicity. Here a distinction is made between the western Hellenic Arc and the Cretan portion of the arc. The difference in b-values between these zones would support such delineation, and the distribution of hypocentres indicates that the western Hellenic Arc is significantly less active than the central and eastern sections.

PZ1990 maintains the distinction between the western [southern Greece], central [Cretan] and eastern [Dodecanese] sections of the Hellenic arc. Each of these sections is given a zone number, yet a further division of each zone is also visible. The western and central Hellenic Arc zones (3, 3A, 4, 4A) are divided into inner and outer arc sub-zones. Division is made on the basis that the catalogue quality is poorer in the outer arc (offshore) than the inner arc. Be it by intention or accident, this division corresponds to the change in focal mechanism from pure thrust in the outer arc to mixed thrust and normal faulting in the inner arc. These reasons are also used as justification for dividing the Dodecanese region into four separate sub-zones (5a, 5b, 5A, 5B). An outer zone (5A) is defined on the basis of catalogue completeness, zones 5a and 5b define the plate boundary and display mixed thrust and normal faulting (though the reason for separating this into two zones is unclear), whilst zone 5B represents a region of exclusively normal faulting. The SW-Turkey/Cyclades band of seismicity (hereafter referred to as the Cyclades-Turkey Extensional Zone, CTEZ) is also included in the PZ1990 zone model (zones 9a and 9b), with a division into two sub-zones made on the basis of orientation of seismicity. The b-values for the western Hellenic Arc region have been designated as 1.0, and for the Dodecanese and normal faulting band 0.8.

The shape of the seismic source zones in the PP2000 is not dissimilar to that of previous models. The Hellenic Arc is divided into a large number of small zones. The western outer arc is divided into four zones, all of which have b-values in the range 0.97 - 1.0. The central section of the outer arc is also divided into four sections, again with b-values in the range 0.98 - 1.01. The western and central sections of the inner arc, which display normal faulting, are divided into five larger zones, running from the Peloponnese to northeast Crete. Again, b-values for these zones are similar, all falling in the range 0.95 - 0.98, which is slightly lower than for the outer arc. As with other areas of the Aegean, the reasons for defining such a large number of zones with similar properties remain unclear. For so many neighbouring zones the differences in b-value are well within the reasonable error in b-value calculation and strike of the major axis of isoseismals differ by only a few degrees. The only significant difference between many zones is the variation in M_{MAX}, yet this may be heavily influenced by the observed seismicity. For 40 of the 67 shallow seismic source zones in the PP2000 model the M_{MAX} defined by the authors has been equalled or exceeded in the historical catalogue.

In the eastern Hellenic Arc the same pattern is repeated, with separate zones defined for the inner and outer sides of the arc. Where there is a notable difference between this model and others is that the junction between the Hellenic and Cypriot arc is placed in a single zone, with a very different orientation to those in the Hellenic arc. In the Dodecanese region b-values are generally lower than for the western Hellenic arc. This is due to the large number of strong historical earthquakes around the island of Rhodos. There are a variety of possible explanations for this difference. The first is that Rhodos and western Turkey have been populated for a long time, and that historical records of large earthquakes are of particularly good quality compared to that of the offshore

sections of the Hellenic arc. However, it is also necessary to consider the higher rates of slip that are observed in the eastern arc compared to the western end, which may result in more frequent powerful earthquakes.

The size and variation in seismic activity rate make the development of a seismic source model along the arc a rather taxing task. As has been shown in previous models, attempts to divide the region into large numbers of small source zones results in neighbouring zones having similar (if not identical) properties. Differences in these zones may be highly dependent on the observed seismicity, particularly M_{MAX}. It is not certain that differences in the zones are representative of differences in true seismicity, or simply artefacts of the brief record of observed seismicity. It may not be prudent to separate a region into smaller zones (with very similar b-values and focal mechanism) simply on the basis that one zone has experienced a large earthquake whilst the other has not. Since large earthquakes may recur on timescales approaching and often exceeding the length of the observed catalogue, it cannot be assumed that this observed behaviour is representative of the true behaviour over the timescales required for design. Hence, in many cases it is more prudent to define just the one zone and allow for a higher uncertainty in the seismic parameters.

A new source model for the Hellenic Arc region is proposed (Figure A.15). On the basis of geodetic deformation and focal mechanism, three shallow zones are developed. The first zone covers the entire outer arc from the southern Ionian Sea to southern Turkey. This zone defines the region of compression and is characterised by low angle thrust faulting with typical strikes of $120^\circ - 150^\circ$. The second zone runs parallel to the first, with the western boundary in southern Greece and the eastern boundary just north of Rhodos. This zone represents the region of east-west extensional faulting, with typical strikes of $-10^\circ - 20^\circ$. A third zone is also defined to characterise the CTEZ. Typical strikes for faults in this region are around 20°.

If only considering focal mechanism and deformation, the zonation proposed here is generally consistent with observations (Figure A.14). Despite this consistency, the distribution of earthquakes and the small difference in GPS movement mean that for both the inner and outer arc zones seismicity is not uniform from east to west. Observed seismicity is substantially lower in the western outer Hellenic Arc and south-western Aegean Sea region, than that in the east. This produces lower b-values in the east of the arc than in the west. This presents a quandary as to whether to assume this differential in seismicity rate is maintained over long timescales, or is a transient feature of the observed catalogue. The fact that a different rate of deformation is observed across the arc would lend some credence to the former. Given the spacing between reliable GPS data points across the region, division of the arc into east and west sub-zones must be done on the basis of observed seismicity.

A division of the two arc zones is made intuitively. An obvious spatial gap in seismicity

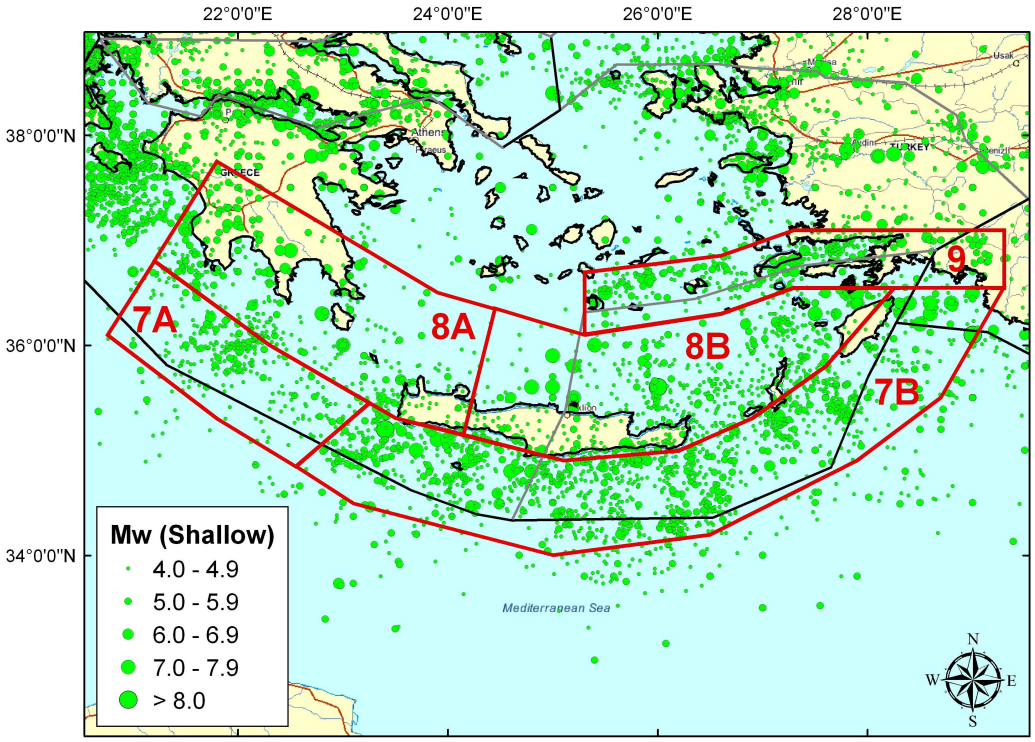

Figure A.15: Proposed shallow source zonation for the Hellenic Arc

of the inner arc is visible in the Cretan sea, to the northwest of Crete. To the west of this zone seismicity is closely related to the extensional faulting of the southern Peloponnese, and to the east the influence of the eastern Hellenic Arc is prevalent. A similar gap exists along the outer arc in the Mediterranean Sea between the south coast of Greece and the western coast of Crete. A NE-SW trending boundary is placed in both the inner and outer arc zones.

In the outer arc, division is also made on the basis of epicentral distribution. It is undesirable to include the cluster of historical powerful earthquakes off the coast of western Crete within the same zone as the seismicity of the Peloponnese. Given the consistency of focal mechanism, however, the inclusion of clusters of smaller seismicity from western Crete into this zone is tolerable. The dividing line of the east and west zones is placed immediately to the west of the epicentre of the M_W 8.3 earthquake. This earthquake is at the western end of a band of high seismic activity that extends across the Hellenic arc to the Dodecanese islands. Unlike PZ1990, variation in catalogue quality in this study is not a primary basis upon which to distinguish between seismic zones. Although previous authors (PZ1990 and PP2000) have decided to split the (CTEZ) into two sub-zones, the variation in activity rate and focal mechanism is not sufficient to define a partition here. As such, this region is kept as one zone.

The b-values for the entire outer arc (zone 7) are 0.760 ± 0.019 (least squares) and 0.784 ± 0.035 (maximum likelihood). For the western sub-division (zone 7A), the b-

values are 1.161 ± 0.098 (least squares) and 1.509 ± 0.190 (maximum likelihood), and for the eastern sub-division (zone 7B) 0.740 ± 0.019 (least square) and 0.768 ± 0.038 (maximum likelihood). If this is compared this previous estimates, zone 7A is spatially similar to HZ1985's zone 3, which is awarded a b-value of 0.88. Zone 7B would represent a combination of HZ1985's zones 4 and 5, which have b-values of 1.23 and 0.90 respectively. The zones of PZ1990 correspond to a similar division, with the western and central arc (zones 3 and 4) having b-values of 1.0, and the eastern arc having a b-value of 0.8. The b-values quoted in PP2000 are different still, with zones in the western and central outer arc regions appearing to have b-values in the range 0.98 - 1.01, compared to 0.90 - 0.95 in the eastern outer arc.

In the inner arc region, more differences are expected. The whole inner arc (zone 8) has b-values of 0.72 ± 0.031 (least squares) and 0.789 ± 0.065 (maximum likelihood). Here the western section (zone 8A) has b-values of 0.804 ± 0.043 (least squares) and 0.954 ± 0.111 (maximum likelihood), and the eastern section (zone 8B) 0.60 ± 0.036 (least squares) and 0.673 ± 0.102 (maximum likelihood). The only correspondence in zones between this model and the HZ1985 model is in the western section of the inner arc. The b-value suggested by HZ1985 for the south Peloponnese region is 1.13. In the PZ1990 model the inner arc is not described by a uniform zone, so comparison cannot be drawn. PP2000 model the inner arc with six zones, whose b-values all lie within the 0.95 - 0.98 ranges.

A brief comparison of b-value is also made for the CTEZ (zone 9). By modelling this region as a single zone the resulting b-values are 0.762 ± 0.062 (least squares) and 1.021 ± 0.133 (maximum likelihood). For this zone, HZ1985 give a b-value of 0.79, PZ1990 assign a b-value of 0.80, and PP2000 divide into three small zones with b-values in the range 0.89 - 0.93 (lower values in the east).

When the least squares b-values are calculated some similarities between source models are visible. There is a general trend of higher b-values in the west of the arc than in the east. Some of this can be attributed to variation in catalogue completeness, but as higher slip rates are observed in the east a physical cause cannot be ruled out. In the low seismicity zone of the western outer arc (7A), there is an appreciable difference between the b-value calculated here and that of HZ1985. This may be due to a swarm of small-moderate earthquakes in this region that occurred during the spring of 1999, in addition to more small earthquakes, which are visible in the earthquake catalogue. There is slight decrease in b-value along the eastern end of the outer arc, but disparity in b-value between the central and eastern zones (4 and 5) in HZ1985 makes comparison largely meaningless. Because there are no comparable zones for the inner arc in the HZ1985 and PZ1990 model, comparison for the inner arc is also spurious. The b-values calculated for the CTEZ (zone 9) are very close to those of HZ1985 and PZ1990 for this region.

The Hellenic arc is the only area of the Aegean region where intermediate depth seismicity (60 - 180 km) is sufficiently spatially coherent to allow for definition of a source zone.

Analyses of focal mechanisms for earthquakes at this depth indicate a large degree of variability, with most focal mechanisms displaying a significant strike-slip component. This is due to along-arc shortening (Benetatos *et al.*, 2004). Although the impact of these events on seismic hazard is not necessarily trivial, the variation in focal mechanism is not sufficient to suggest that large differences in ground motion due to directivity effects may be observed. Previous models of intermediate depth seismicity in PZ1990 and PP2000 have tended to define zones along two arcs, with depths corresponding to the location of the Benioff zone of the subducting African slab. Given the brevity of the time period for which accurate depth location has been possible, there are very few events from which estimates of the Gutenberg-Richter parameters can be made. Delineation of zones of intermediate depth seismicity is largely dependent on the spatial clustering of seismicity. Although a down-dipping slab is visible from observed hypocentres, intermediate depth seismicity is still well dispersed. It is still necessary, therefore to define source zones of uniform depth rather than a volumetric source for a subducting slab.

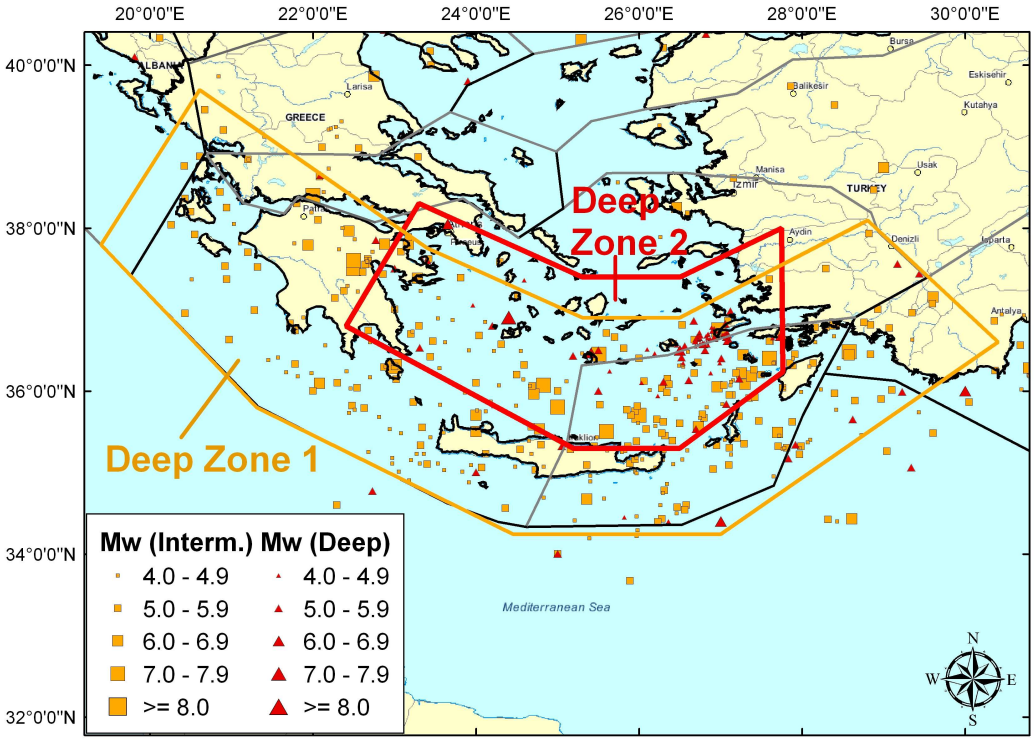

Figure A.16: Proposed seismic source zonation for intermediate (depth 60 120 km [Zone 1]) and deep (depth 120 180 km [Zone 2]) earthquakes

A proposed source model for intermediate depth earthquakes is shown in Figure A.16. In this model two zones are defined, each of constant depth and uniform seismicity. Zone D1 is the intermediate zone, which is the source for earthquakes with depths in the range of 60 - 100 km. Zone D2 is the source for earthquakes in the depth range 100 - 180 km. Given the small number of earthquakes under consideration, unlike the PP2000 model, it is not necessary to further divide these regions into west, central and eastern components. However, the shapes of these two zones are consistent with previous source models and

reflect the shape and location of the down-dipping slab.

As there are so few earthquakes to accurately determine b-values for intermediate depths, estimates carry large uncertainties. Both PZ1990 and PP2000 use the same b-value for the outer (0.56) and inner arc (0.75) intermediate depth earthquakes. In comparison, zones D1 and D2 have b-values of 0.61 ± 0.024 and 0.870 ± 0.054 (both using the least squares method) respectively. These values are certainly consistent with those estimated previously.

A.6 Western Turkey, the North Aegean Sea and the North Anatolian Fault

So far in this analysis, the Aegean region has mostly been considered in terms of a typical continental collision and subduction environment. This assertion is supported by the abundant thrust faulting along the Adriatic coast (orogenic collision) and Hellenic arc (subduction), the strike-slip faulting in the Ionian islands, and the normal faulting and lithospheric extension that is seen in much of Greece, the south Aegean Sea and south west western Turkey (back-arc extension). The Aegean region is particularly complex compared to many subduction zones, however, due to the influence of the North Anatolian Fault (NAF). The NAF defines the northern boundary of the Anatolian plate, and is manifest as a single fault trace extending 1000km from Karliova, Eastern Turkey, to Mudurnu, NW Turkey (Gürer *et al.*, 2005). The situation becomes more complex to the west of Mudurnu (at the very easterly extent of this study region), as it divides into separate branches (Figure A.17). The geomorphology of much of NW Turkey is representative of this division, with many rhomb-like pull-apart basins (Armijo *et al.*, 2002).

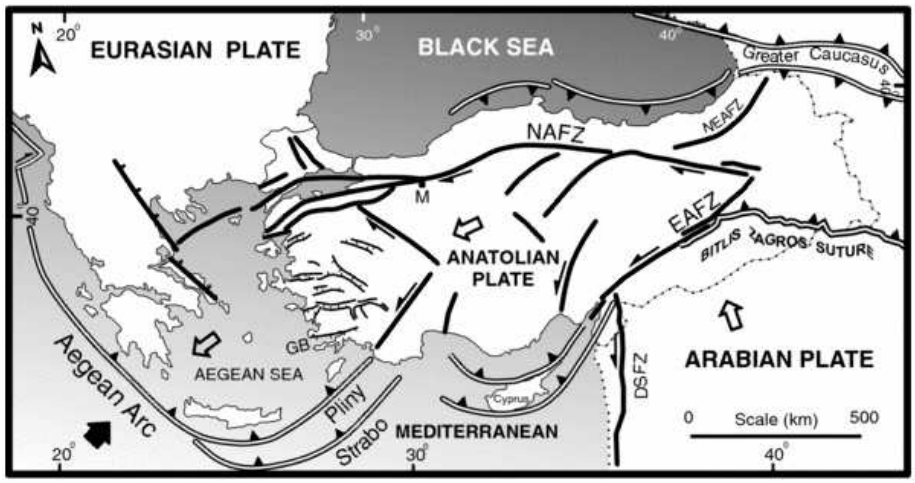

Figure A.17: Tectonics of the Anatolian Plate (as interpreted by Gürer *et al.* (2005)). M is the town of Mudurnu, East Anatolian Fault Zone (EAFZ), Dead Sea Fault Zone (DSFZ)

Along the single fault trace, the movement of the fault is largely dextral strike-slip. Estimates of the rate of strike-slip motion for the NAF along much of this trace seem to

suggest approximately 25 - 30 mm/yr (Reilinger *et al.*, 2000; Papazachos, 1999b; Reilinger *et al.*, 2006). However, where the NAF divides in the NW Turkey region, slip partitioning is observed. Most of the displacement is accommodated by the northern branch of the NAF in the Sea of Marmara. The historical seismicity of the NW Turkey region clearly indicates that the northern branch of the NAF is more active than the central and southern branches. Slip on this section of the NAF is estimated to be in the range 20 - 25 mm/yr (Armijo *et al.*, 2002). Although movement on the northern branch of the NAF is principally strike-slip (Figure A.18), there is abundant evidence from the focal mechanism and bathymetry of the Sea of Marmara for normal faulting in minor basins within the Sea of Marmara (Armijo *et al.*, 2002; Le Pichon *et al.*, 2003). Along this fault four small rhomb-like basin structures are visible, which display normal faulting characteristics typical of pull-apart basins.

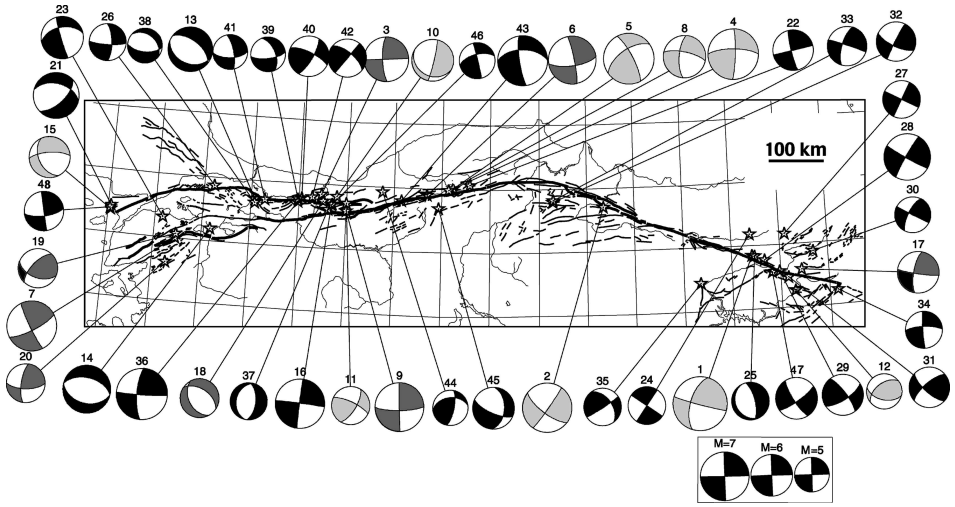

Figure A.18: Focal mechanisms of strong earthquakes along the NAF from 1939-2003 (Image courtesy of Segnor *et al.* (2005))

There is some controversy regarding the origin of the Sea of Marmara. Specifically, whether it is a pull-apart basin arising from the current dextral strike-slip regime (Armijo *et al.*, 2002), or whether it is exhibiting dextral strike-slip movement on an older extensional basin (Le Pichon *et al.*, 2003). Evidence for the former hypothesis comes from the second order pull-apart basins visible in the ocean bathymetry. Le Pichon *et al.* (2003) refute this, citing the absence of lateral steps along the sides of these smaller basins. They point to evidence of dextral offsets within these basins, which suggest that the Northern Marmara Fault System (NMFS) behaves like a dextral strike-slip fault striking through an old extensional basin. Seismic reflection data indicates that the trace of a coherent fault underlying the small basins within the NMFS is visible, which would not necessarily be apparent in a pull-apart basin. It is still not clear from either of these hypotheses why the single fault trace of the NAF partitions to begin with. The extensional tectonics arising from the back-arc of the Hellenic subduction zone obviously influences the faulting. This is seen more clearly in the North Aegean Trough (NAT) (McNeill *et al.*, 2004).

In the eastern Sea of Marmara the NMFS turns from E-W to NW-SE, parallel to the coast-

line, before turning back to E-W strike faulting along the northern shelf of the basin. Along the NW-SE striking section focal mechanisms display a normal-oblique component of faulting. This is due to the offset between the direction of strike and the direction of principal stress. It is this extensional faulting that gives rise to the Çinarcik Basin in the eastern Sea of Marmara. Along this section of the NMFS there have been few instrumentally recorded earthquakes from which accurate focal mechanisms can be determined. Kiratzi (2002) identifies two focal mechanisms along the Çinarcik fault, both of which show normal faulting, with the active plane striking at 305°. Pinar *et al.* (2003) have determined focal mechanisms displaying predominantly strike-slip behaviour, however. They attribute normal faulting to the Yalova-Hersek fault, found onshore on the Yalova peninsula. This distinction is crucial in trying to understand the tectonics of the Sea of Marmara and asses the seismic hazard to Istanbul. If the Çinarcik fault is undergoing dextral strike-slip faulting then this would support the assertion that the Marmara basin is not a pull-apart (Le Pichon *et al.*, 2003), but was initially formed by extension, and is now subject to E-W strike-slip faulting. Pinar *et al.* (2003) propose a seismotectonic model for this region whereby the Çinarcik fault is a step-over between the western extension of the single-trace NAF and the Northern Marmara fault. This may explain some of the seismicity observed in the Istanbul Metropolitan area in terms of antithetic faults.

The E-W striking Northern Marmara (or Central Basin) fault is of particular concern for seismic hazard in the region. Much of this fault has been seismically quiescent within the last 500 years. The largest earthquake to occur along the eastern section of this fault was the 1509 Istanbul earthquake ($M_W \approx 7.2$). Between this segment and the rupture of the 1912 Saros-Marmara earthquake there has been little observed activity. Ambraseys and Jackson (2000) speculate on the possibility of a seismic gap in this particular segment.

To the immediate west of the Sea of Marmara the E-W striking Marmara fault turns to the southwest to become the Ganos fault. This fault produces some varied morphology in western Turkey. Just east of the bend, there is another small pull-apart structure in the form of the Tekirdag Basin, yet immediately to the west there is a region of compression, manifest in Ganos Mountain. Okay *et al.* (2004) consider this an example of the transition from transpression to transtension. The fault then crosses the Gallipoli peninsula, emerging into the Gulf of Saros in the northeast Aegean Sea. This fault is known to be active, having ruptured in the 9 August 1912 (M_W 7.1) Saros-Marmara earthquake.

Due to the potential seismic threat to the city of Istanbul, the time-dependent behaviour of earthquakes in the Sea of Marmara has been under scrutiny since the 1999 Izmit earthquake. The quality of the Turkish earthquake catalogue, and the clear traces of ruptures along the NAF, has made this particular region an ideal case study for the investigation of time-dependent earthquake behaviour. Models of coulomb stress changes for earthquakes preceding the Izmit earthquake show a clear pattern of rupture along segments where coulomb stress is raised by a preceding earthquake (Figure A.19).

The Coulomb stress change visible in Figure A.19 is the western extension of a trend of westward progressing ruptures along the entire length of the NAF in the 20th century. It is clear that the Izmit earthquake has created a substantial increase in Coulomb stress along the Cinarcik and Southern Boundary faults. Rupture along the Çinarcik fault in 1766 (M_W 7.2) produced MMI values as high as VIII or IX in Istanbul (Parsons, 2004). Similarly, Coulomb stress change along the western and central sections of the NMFS, following the 1912 Saros-Marmara earthquake and 1935 Marmara Island events (two M_W = 6 events), has increased the probability of rupture along the segment that ruptured during the 1509 earthquake (M_W 7.2, MMI at Istanbul = IX). Along the Marmara and Gulf of Saros sections of the NAF, seismicity cannot be considered independent over short timescales. This is an issue to consider in the zoning scheme.

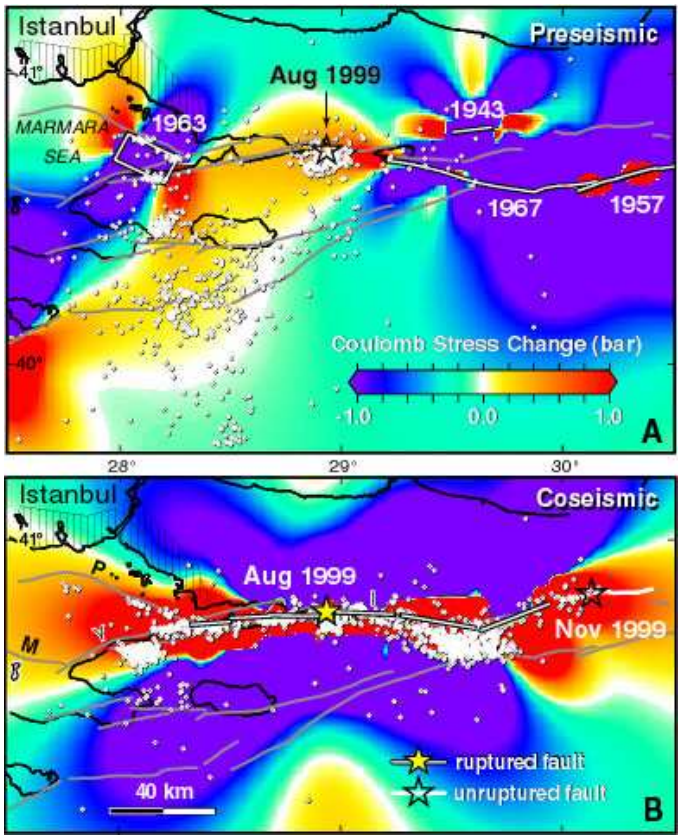

Figure A.19: Coulomb stress change preceding (upper) and succeeding (lower) the August 1999 Izmit and November 1999 Duzce earthquake (Parsons *et al.*, 2000)

The surface rupture arising from the 1912 Saros-Marmara earthquake provides a clear indication of the extension of the NAF into the North Aegean Sea. The delineation of the fault becomes less well-defined further westwards. From the bathymetry of the North Aegean Sea a trough is visible, which follows the expected line of the northern branch of the NAF. This trough extends from the Gulf of Saros to the Sporades Basin, terminating off the coast of Thessalia (Eastern Greece). The morphology of this basin reveals major strike slip and normal faults running in a NE-SW direction. In addition, numerous minor faults, oblique to the main direction of strike, have been identified by seismic reflection profiling (McNeill *et al.*, 2004). These minor faults are indicative of Reidel shears, which

would be expected in the strike-slip faulting environment.

Whilst strike-slip faulting is evident, the role of extensional tectonics cannot be ignored. Geophysical investigation of the North Aegean region indicates that the thinnest lithosphere is found beneath the North Aegean Trough (NAT) (Le Pichon and Angelier, 1981; Makropoulos and Burton, 1984; Kalogeras and Burton, 1996). This would strongly suggest that some extensional process operates, and that the NAT is not simply a pull-apart basin.

The model proposed by McNeill *et al.* (2004) to account for these observations is a strain-partitioning model. In the Gulf of Saros the trace of the NAF is a reasonable approximation of a dextral strike-slip fault, as seen along the rupture of the 1912 earthquake. In the Saros basin, the fault partitions into three sections at shallow depths. The outer branches have a significant oblique extensional component and mark the edges of the trough. The trough itself is bisected by the central partition, which accommodates most of the dextral strike slip motion. Shallow seismic reflection data would suggest a limiting depth of approximately 2km, beneath which a single dextral strike-slip fault may be defined.

Focal mechanisms for the NAT indicate strike-slip faulting along a fault plane striking in a NE-SW direction. This puts the principal axis of tension in a north-south direction, in line with the principal axes of stress for extensional faults observed across most of mainland Greece and western Turkey (Kiratzi, 2002). Some focal mechanisms display an oblique component, which is usually further indicative of extension in the NAT (Kiratzi and Louvari, 2003). In accordance with the change in strike from ENE-WSW to NE-SW across the NAT, the fault planes of the focal mechanisms change alignment.

Seismicity in the NAT is less active than that observed further east in the Sea of Marmara. In spite of this, more then 10 large earthquakes ($M_W \geq 7$) have been recorded in this region historically, many of them associated with destructive tsunamis in the region. GPS data suggest that movement in the North Aegean is not significantly different from that in NW Turkey (Nyst and Thatcher, 2004). Likewise, resultant slip rates predicted by the REIL2006 block model indicate that slip is mostly constant along the northern branch of the NAF, as far as the Sporades basin. The only difference between the nature of slip in the western and eastern NAF in the NAT is that, in the western NAT, some dextral strike-slip faulting gives way to extensional slip. The resultant slip is still a constant 26 - 27 mm/yr.

There is some disparity in previous tectonic models of the Aegean regarding the western termination of the NAF. McKenzie (1978), Nyst and Thatcher (2004) and Reilinger *et al.* (2006) describe block models that require the delineation of a connecting fault between the NAF in the Sporades basin and the northern CTF. Whilst this may be inferred from a geodesy, the seismicity of central Greece does not indicate any such fault (Goldsworthy *et al.*, 2002). In section A.2, the North Greece/South Balkan Extensional Zone was

described as a region of extensional faulting with seismicity occurring in tight clusters. The south-eastern boundary of this zone was placed between the earthquake cluster observed in the Gulf of Volos, and one observed in the Sporades basin. Across this boundary there is a change in focal mechanism from dextral strike-slip (or strike-slip containing an oblique extensional component) faulting to normal faulting. From a purely seismic perspective it is reasonable to define the Sporades region as the western termination of the northern branch of the NAF.

Whilst it may be the northern branch of the NAF that accommodates most of the slip, and presents the greatest hazard to major cities in the Marmara region, the central and southern branches cannot be excluded from consideration. Although fewer powerful earthquakes have been recorded along this branch compared to the northern branch, nine strong earthquakes ($M_W \geq 7.0$) are found in the historical catalogue. Along this southern branch the delineation of a single coherent fault system is less clear. It is apparent that active faults in this region are more segmented over long distances. Furthermore, the faults along the southern branches of the NAF have a greater extensional component than those of the northern branch. This is visible in the geomorphology of the southern Marmara region, with many lakes and depressions marking the trace of the smaller faults (Figure A.20 - Gürer *et al.* (2005)). Although the broader tectonic regime of the south Marmara region is largely extensional, a number of small reverse faults are also visible. The strikes of the fault planes of these reverse faults are similar to those of the normal faults, suggesting that these are antithetic faults arising from the normal faulting. Focal mechanisms along the southern Sea of Marmara clearly indicate dextral strike-slip faulting with an oblique normal component. The variation in strike of the fault planes is a result of variation in fault strike and local stress perturbations (Gürer *et al.*, 2005).

The absence of a single fault structure accounting for the dextral strike-slip motion in the southern Marmara region is an important factor in delineating seismic source zones. Figure A.20 clearly illustrates the complexity of the fault network in this region. Given the proximity of neighbouring faults it is unrealistic to assume that activity on these faults is independent. In this respect there is very little reason to attempt to define separate zones for the central and southern branches of the NAF or smaller faults.

The eastern termination of the southern Marmara fault system is the junction where the NAF partitions (west of Izmit), yet the western termination is less obvious. As with the northern branch of the fault, focal mechanisms suggest that the dextral strike-slip faulting from these branches extends into the Aegean Sea. There is some evidence in the bathymetry in the form of another smaller trough in the central Aegean, which runs parallel to the NAT. This is the Skyros-Edremit trough. It is not as deep as the NAT, which is consistent with the smaller rate of deformation.

The spatial distribution of strike-slip focal mechanisms appears to extend as far as the island of Evia, with the extensional contribution increasing from east to west (Kiratzi,

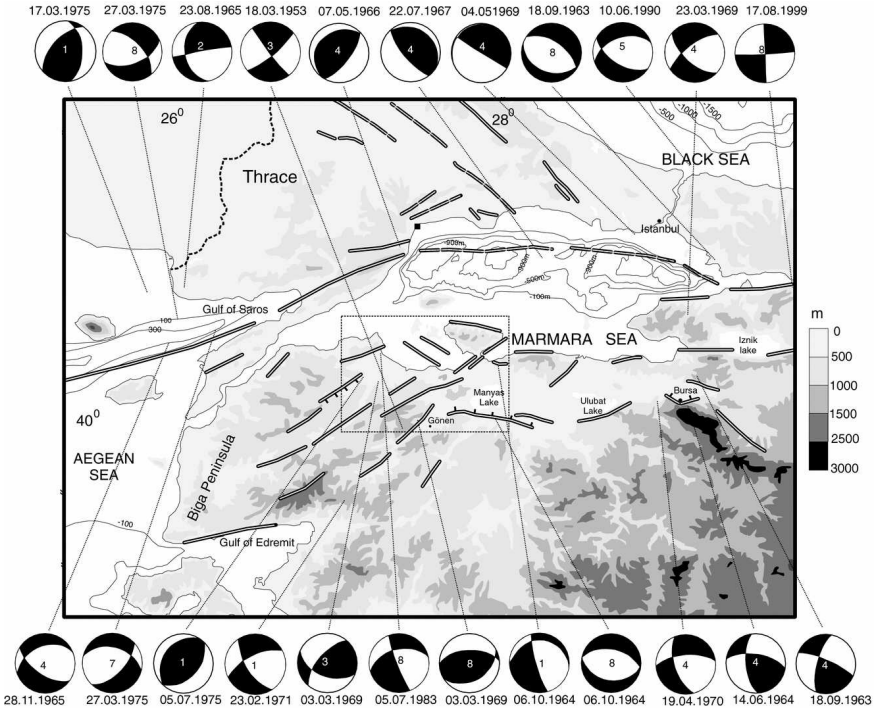

Figure A.20: Surface trace of active faults in the Marmara region with fault plane solutions marked (Gürer *et al.*, 2005)

2002). Focal mechanisms from the 2001 Skyros earthquake (M_W 6.4) and its aftershocks indicate that strike-slip faulting is still dominant to the immediate west of Skyros (Ganas *et al.*, 2005). For this earthquake, the fault plane is believed to strike from NW-SE. This is orthogonal to the strike of the NAF in this region. During the Skyros earthquake, sinistral slip was observed, in opposition to the dextral slip expected across this region. The axes of tension and compression still remain the same, however, indicating that this earthquake is still consistent with the regional direction of stress.

GPS data suggest that along the southern branches of the NAF, movement (relative to a fixed Eurasia frame) is similar to that of the northern branch of the NAF (Nyst and Thatcher, 2004). Since most of this movement is accommodated along the northern branch then slip rates are lower along the southern branch. This is visible in the REIL2006 block model, which estimates rates of 5 - 7 mm/yr along the southern branch. Most of this is accommodated by extension, which is not entirely consistent with the style of faulting.

Although the two dominating features of seismicity in western Turkey are the north and south branches of the NAF, much of western Turkey between the NAF and the Hellenic Arc is still highly active. Many major cities in this region have suffered damage from powerful earthquakes in recorded history, the most recent being the 28 March 1969 Alasehir (M_W = 6.6) earthquake and 28 March 1970 Gediz (M_W = 7.0) earthquake (Zanchi and Angelier, 1993). As with much of the central Aegean, the tectonic regime is predominantly extensional faulting. Geomorphologic evidence for this can be found in the

numerous graben structures across western Turkey. Analyses of focal mechanisms by Zanchi and Angelier (1993) also confirm this trend of NE-SW extension. This tectonic pattern is geodynamically consistent with the model of an anti-clockwise rotating Anatolian plate overriding a subducting slab in the Hellenic arc. The extension arises because of the arc-normal pull from the Hellenic trench, in conjunction with westward push due to the greater lithospheric thickness in eastern Turkey (Meijer and Wortel, 1997).

GPS velocities in this region are similar to those found in the central Aegean (Nyst and Thatcher, 2004). The geodynamic interpretation of this feature appears to vary. Both Nyst and Thatcher (2004) and Reilinger *et al.* (2006) attribute the extension in Western Turkey to the boundary of a South Aegean microplate. Earlier interpretations (Goldsworthy *et al.*, 2002) define the Anatolian and South Aegean as a single block, which exhibits internal deformation and extension. Whilst Nyst and Thatcher (2004) do recognise that if a plate boundary between the South Aegean and Anatolian microplates does exist, it is poorly constrained by the observed seismicity. In the REIL2006 block model, an attempt to define the boundary as a coherent fault structure is made. In doing so, they define a line of faults with extensional slip rates of 14 - 16 mm/yr, and strike-slip rates varying between 1 - 10 mm/yr (dextral or sinistral) depending on fault strike.

Before creating a new source model for western Turkey, it is interesting to see how this region has been modelled in previous zonations. One of the greatest concerns with existing source models has been the assumptions regarding seismicity in western Turkey. The HZ1985 model separates western Anatolia and the Aegean Sea into three E-W trending rectangular zones (zones 12, 13 and 14) with no western boundary. The trace of the NAF is divided into a western zone (zone 17, the NAT) and an eastern zone (zone 20, the Sea of Marmara). The choice of simple rectangles to model the seismicity of a complex fault structure has a significant influence on seismic hazard assessment. The b-values for this zone decrease from SW to NE. The b-values for the four zones in western Turkey and the central Aegean Sea lie in the range 0.78 - 0.92. The NAT has a b-value of 0.62, and the Sea of Marmara 0.56.

The PZ1990 model has many similar features as the HZ1985 model in this region. Again, the decision has been made to divide western Anatolia into three parallel rectangles (zone 12, 13 and 14), this time with a well-defined western terminus. Most of the central and south Aegean Sea is left as background. The NAT (zone 16) and Sea of Marmara (zone 15) zones remain identical to those of HZ1985. The most significant difference between the PZ90 and HZ1985 model is in the north Aegean Sea. In the PZ1990 model zone 14 runs east to west from NW Turkey to Thessalia. This is separated into four sub-zones: Western Turkey (14d), Gulf of Edremit (14c), Skyros (14b) and Sporades (14a). Papazachos (1990) states that the reasons for separating central Western Turkey into two zones (zones 12 and 13), despite having the same seismicity and fault type, are based on geomorphologic features, but elaborates no further!

The reasons for dividing zone 14 into four sub-zones arise largely from differences in focal mechanism. The model recognises that the western (14a and 14b) and eastern (14c and 14d) sections of this zone must be separated, due to the prevalence of strike slip faulting in the west and normal faulting in the east. Zones 14a and 14b are separated on the basis that two bands of parallel strike-slip faulting are observed, zone 14a corresponding to the northern branch of the NAF, zone 14b the central and southern branches. Zones 14c and 14d are separated on the basis of data quality with 14c having more historical data. A vague justification of "geometrical reasons" is given for separating the NAT from the Sea of Marmara. Corresponding with the general trend of diminishing b-value to the northeast, zones 12 - 14 are assigned b-values of 0.8, whilst zones 15 and 16 are assigned b-values of 0.6.

Many of the same features of the PZ1990 model are maintained in the PP2000 model, albeit with further sub-division. The principal difference is that the central and southern Aegean Sea is now clearly divided into closed source zones, thus allowing for the possibility of an earthquake anywhere in the Aegean region. Parallel rectangular zones are still defined in western Anatolia (zones 53, 54, 56-58, 62 and 63). The division of zone 14 in the PZ1990 model is kept the same in the PP2000 model, the only difference being a further division of 14d into two zones (62 and 63). This further division may be based on differences in catalogue quality.

In the PP2000 model, the trace of the northern branch of the NAF is almost entirely lost. The Sea of Marmara is divided into an east and west section, yet the southern boundary of these zones have been extended further south to include the southern branch of the NAF. The NAT is also divided into two zones with the Gulf of Saros being separated from the Chalkidiki peninsular region. Again, there is very little explanation for these particular subdivisions. The b-values for the western Turkey region all lay within the range 0.84 - 0.88, whereas for the zones closely related to the NAF the b-values are mostly within the range 0.80 - 0.84. This lack of variation is, again, most likely due to the b-value smoothing approach implemented by the authors. The western Turkey zones (52 - 58) seem to have major axes of isoseismals striking between 50 and 80 , which corresponds to the WSE-ENE striking extensional faulting.

One further zonation scheme that is pertinent to seismic hazard in western Turkey and the Aegean is that of Erdik *et al.* (1999). In many respects this zonation is the closest to the observed tectonic structures in the region. The key distinguishing feature of this model is the delineation of two elongated fault zones that follow the traces of the northern and southern branches of the NAF from the division point, west of Izmit, to the Sporades basin. These narrow zones allow for uncertainty in the location of earthquakes attributed to the NAF, and for the partitioning of the faults as they enter the Aegean Sea. They also follow closely the line of dextral strike-slip faults observed in the focal mechanisms. A separation of the northern NAF between the Sea of Marmara and NAT is made, although the dividing line crosses the rupture of the 1912 earthquake and, hence, may not

be entirely justified. The source zonation scheme in western Turkey corresponds to the outlines of the largest graben structures in the region. These zones correlate to the highest seismicity in the region, yet they may be somewhat arbitrary as they make no allowance for the inter-dependence of these graben systems. Furthermore, seismicity is so spatially diffuse in this region that strong earthquakes are observed outside of these source zones.

Having seen how the NAF and western Turkey regions have been zoned previously, new source models are presented for this region. The new zonation scheme can be seen in Figure A.21. In line with the zonation methodology used elsewhere in this model, a minimalist approach is implemented. The only coherent fault structures in this region stem from the western extension of the NAF. Because of the difference in the amount of strain accommodated, and therefore the parameters of the frequency-magnitude relation, northern and southern branches of the NAF are separated. This is seen in the shape of zones 10 (northern) and 11 (southern).

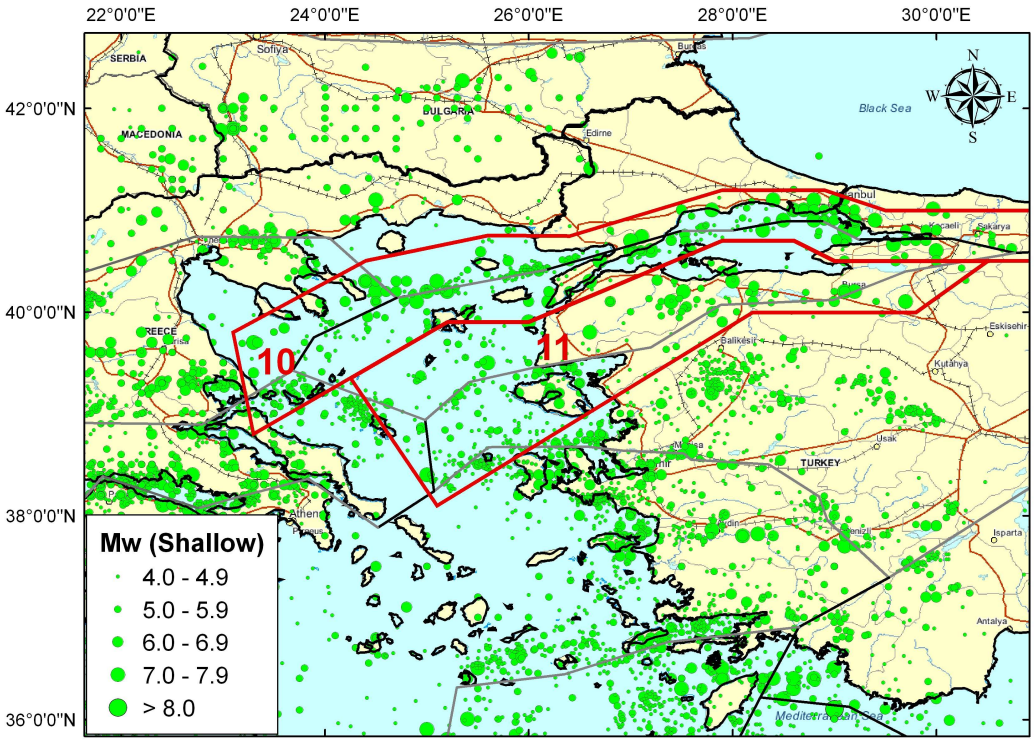

Figure A.21: Proposed shallow source zonation for the North Anatolian Fault in the Aegean region

Zone 10 extends as far east as the city of Izmit, and it follows the trace of the northern branch of the NAF. The western boundary is located such that seismicity in Sporades islands is included, but seismicity in the Gulf of Evia is not. This decision is simply made on the basis of focal mechanisms, as dextral strike-slip faulting is still visible the Sporades region. The width of the zone increases from east to west due to the fault partition in the NAT, and seismicity is increasingly influenced by the pull of the Hellenic Arc. The southern boundary of zone 10 follows a virtual spatial gap in the seismicity. It

extends south to the island of Bozca to incorporate the location of the 1672 earthquake (M_W 7). Although the trace of the northern NAF is visible in the Sea of Marmara, it is still necessary to define a zone of reasonable spatial width to allow for uncertainty in location and activity on branch faults. This can be seen in the Istanbul region.

Zone 11 follows the approximate trace of the southern and central branches of the NAF. Again the western terminus occurs at the city of Izmit and it follows the approximate trace of the fault into the Gulf of Edremit. The eastern terminus is the island of Skyros as it is here that sinistral strike-slip faulting emerges, as indicated by the 2001 earthquake (Ganas *et al.*, 2005). This zone is substantially wider as it incorporates many of the active faults that are observed in the Biga region of western Turkey.

The b-values for both of these zones are lower than the rest of the Aegean region, thus following the trend that has been evident in previous source models. Zone 10 has b-values of 0.658 ± 0.062 (least squares) and 0.458 ± 0.027 (maximum likelihood). Zone 11 has b-values of 0.695 ± 0.062 (least squares) and 0.578 ± 0.096 (maximum likelihood). The completeness magnitude in this region is $M_W \approx 5.2$. This figure has been remarkably consistent on both a region wide scale and a localised scale. Although the shapes of the zones differ significantly from those in the previous models (the exception being the EK1999 model), these b-values are not dissimilar. The reason for the lower b-values is simply that we have a large number of high magnitude earthquakes in the historical record compared to those of smaller earthquakes. It is unclear whether this trend is due to the higher slip rates along these faults or due to the poor record of small earthquakes in the historical catalogue.

The only region left to consider is western Turkey and the South Aegean Sea. In the previous models this region has been characterised by a series of parallel rectangles, or, in the case of the Aegean Sea, left unzoned. This approach gives some cause for concern as the observed seismicity simply doesn't conform to this pattern. The question then arises; what, if any, pattern does the seismicity of western Anatolia conform to? Although there is some variation in focal mechanism, the overwhelming trend is for normal faulting with major axes of tension in a north-south direction. Variation in this pattern is mostly due to differences in strike and dip of faults that are active in this regime. Kinematic models of the region suggest a boundary between a South Aegean microplate and the Anatolian microplate, but a coherent boundary is not evident in the distribution of seismicity. The Erdik *et al.* (1999) approach of defining zones based on the shape of known grabens would seem logical, yet once again the seismicity does not conform to this particular idealisation.

This is not dissimilar to the situation encountered in the South Balkan Extensional Zone, seen in section A.2. There, as in this region, it is clear that seismicity is not uniform, but there is not a sufficiently sound knowledge of active faults to delineate smaller source zones. Given that focal mechanisms and geodetic measurements do not vary across this

zone to such a degree that it would be easy to identify different tectonic environments, the only source of information remaining is the observed seismicity. The immediate answer to this problem is to not zone. Here, as with the Northern Greece/South Balkan region, the entire area is defined as a single non-uniform region of seismicity. That is to say the only indication of future seismicity is the distribution of observed seismicity. More discussion on how these zones can be used in seismic hazard analysis is presented in Chapter 3.

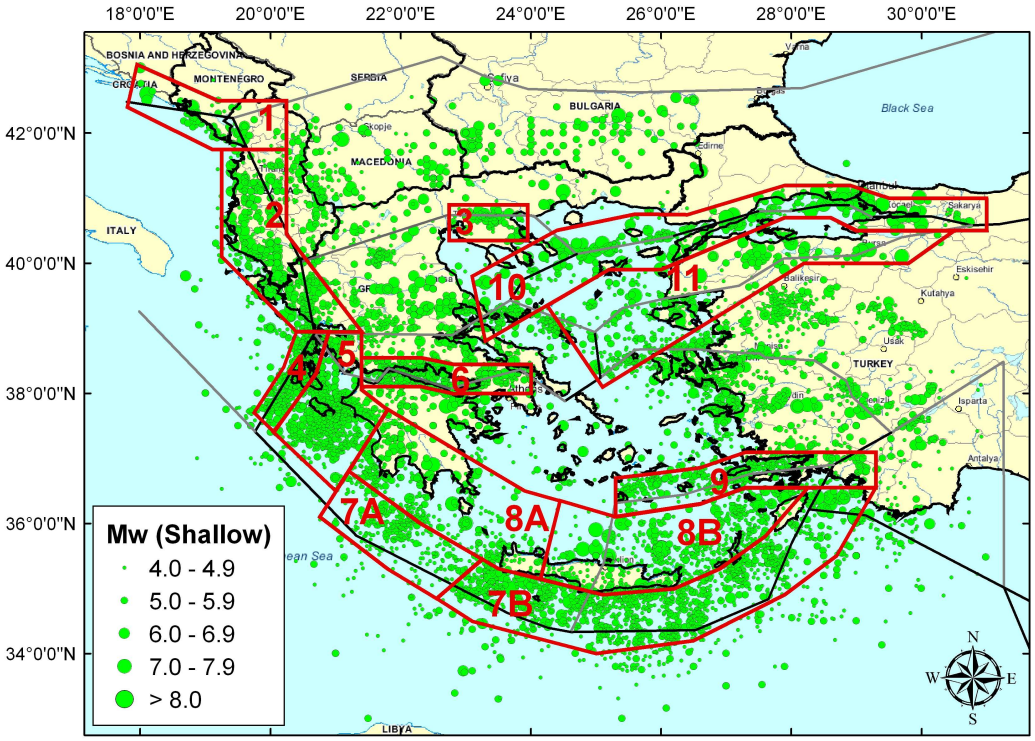

Figure A.22: The WT2006 uniform source model of shallow seismicity for the Aegean region, shown with observed seismicity and the REIL2006 geodynamic model

The final source model is seen in Figure A.22 with b-values given Table 3.1. This model for the delineation of shallow seismic sources represents this author's judgement, given the seismotectonic information available. There are clearly many consistencies with alternative models, and also some areas where this model presents a different interpretation. Other seismologists and geologists may present different interpretations. The arguments for the delineation of each zone presented here provide transparency with regards to the decisions made in its development. This may allow other seismologists to better judge how this model should be viewed or weighted alongside existing and future seismic source models in the Aegean region.

Appendix B

Electronic Appendix - Data Files

1. The Extended Aegean Earthquake Catalogue

   The data file is a combination of the Papazachos and Papazachou (1997) historical earthquake catalogue (pre - 1900), the Burton *et al.* (2004a) instrumental catalogue (1901 - 1999) and the National Observatory of Athens Earthquake Catalogue (2000 - 2005). See Chapter 2 for more details.
   Filename = Appendix2_EarthquakeCatalogues.txt
   Format = ascii (tab delimited)

2. WT2006 Zone Coordinates

   This file contains the coordinates of the WT2006 shallow and deep source zones, described in Chapter 3.
   Filename = Appendix2_WT2006Coordinates.txt
   Format = ascii (tab delimited)

3. List of Strong Ground Motion records used for quantitative fit of attenuation relations (Chapter 5).

   This file contains a list of the strong ground motion records taken from the European Strong Motion Database (Ambraseys *et al.*, 2004) used for quantitative analysis of the fit of attenuation relation to Aegean strong motion data.
   Filename = Appendix2_SGMEarthquakes.txt
   Format = ascii (tab delimited)

4. Parameters of fit of spectral attenuation models to spectral acceleration data (Chapter 5)

   This file contains the Mean of Normalised Residuals (E(Z)), Median of Normalised Residuals (Z_{50}), Standard Deviation of Normalised Residuals (Z_σ), Likelihood Fit (LH_50) and Scherbaum *et al.* (2004a) Class. This data is plotted in Figure 5.16.
   Filename = Appendix2_SGMSpectralLHFit.txt
   Format = ascii (tab delimited)

5. Microsoft Excel file containing all the above databases

   Filename = AppendixDatabases.xls
   Format = Microsoft Excel Spreadsheet (.xls)

6. Seismic Hazard Maps for the 5 % and 2 % probability of being exceeded in 50 years

   Hazard maps for PGA with a 10 % proability of being exceeded in 50 years for given combinations of models are found in Figure 6.4 and Figures 7.4 to 7.7. The equivalent maps for the PGA with a 5 % and 2 % probability of being exceeded in 50 years are given here.

   Filename = Appendix2_LowProbHazardMaps.pdf
   Format = Portable Document Format (.pdf)